RECONOCIMIENTO MOLECULAR BASADO EN TÉCNICAS DE RMN

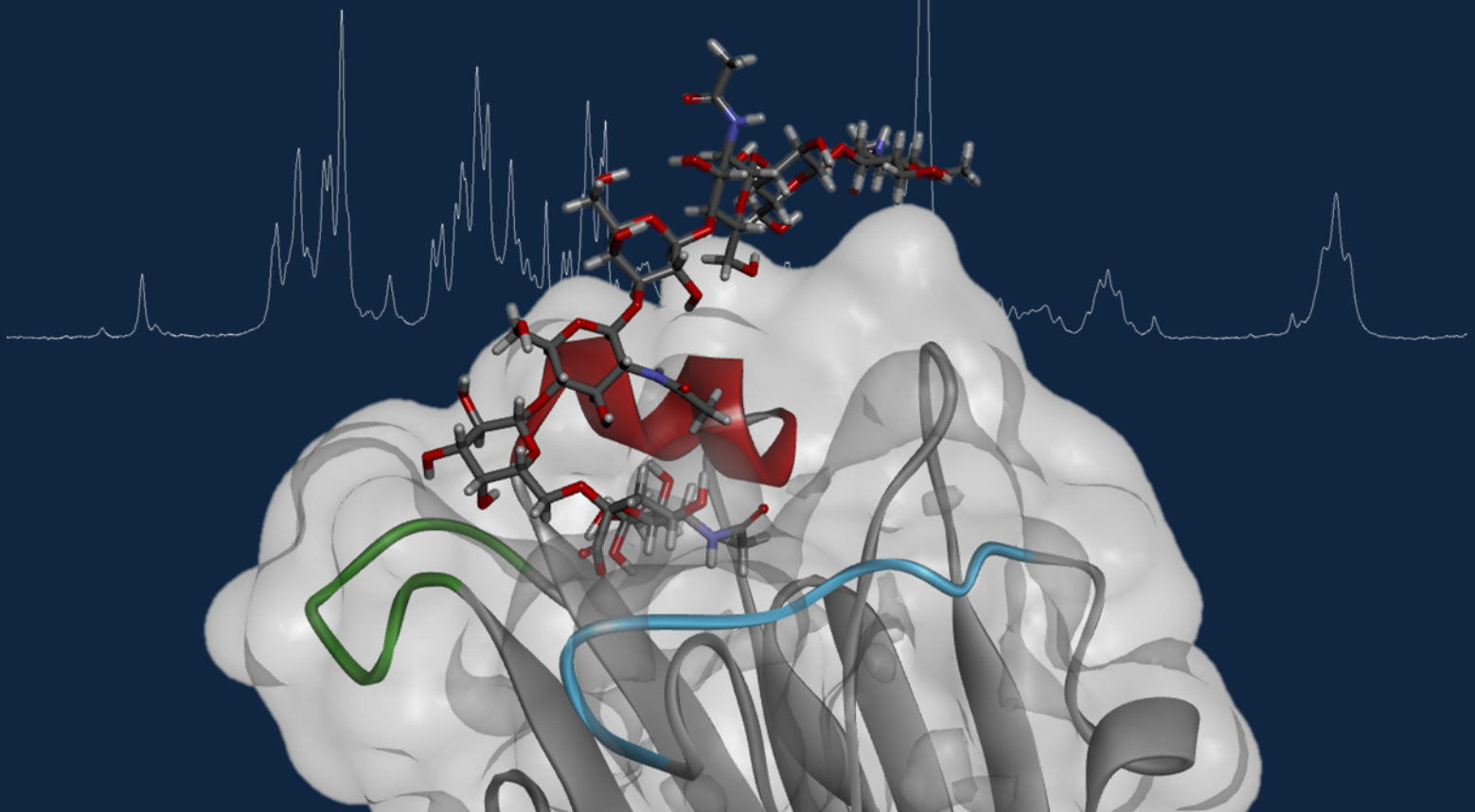
NMR & MOLECULAR RECOGNITION

Facultad de Ciencias Químicas

TESIS DOCTORAL

Javier Sastre Martínez

2017



UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS QUÍMICAS

TESIS DOCTORAL

Reconocimiento molecular basado en técnicas de RMN

NMR & molecular recognition

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Javier Sastre Martínez

Directores

Jesús Jiménez-Barbero
F. Javier Cañada Vicinay

Madrid, 2018

UNIVERSIDAD COMPLUTENSE DE MADRID

Facultad de Ciencias Químicas

RECONOCIMIENTO MOLECULAR BASADO EN TÉCNICAS DE RMN

NMR & MOLECULAR RECOGNITION

Memoria para optar al grado de Doctor presentada por

Javier Sastre Martínez

Madrid, 2017

Esta Tesis ha sido realizada en el Centro de Investigaciones Biológicas

(CIB-CSIC)

Bajo la codirección de

Dr. Jesús Jiménez-Barbero

Dr. F. Javier Cañada Vicinay

Acknowledgements / Agradecimientos

El trabajo presentado en esta tesis ha sido realizado en el centro de investigaciones biológicas del CSIC (CIB-CSIC) gracias a la financiación recibida del ministerio de economía y competitividad (MINECO) para la formación de personal investigador (BES-2012-062847).

En primer lugar, agradecer a mis directores de tesis Jesús Jiménez Barbero y Javier Cañada por darme la oportunidad de haber realizado estos 4 años de doctorado en su laboratorio y haberme sabido transmitir sus conocimientos de RMN aplicados al reconocimiento molecular. Gracias a ellos he podido realizar esta labor de investigación que concluye con la defensa de esta Tesis.

No puedo olvidarme tampoco de todos aquellos compañeros de laboratorio con los que he tenido el privilegio de compartir estos años de formación y que también han contribuido a mi formación personal y científica de numerosas maneras: Ana, Pilar, Khouzaima, Dolores, Anita, Bea, Álvaro, Ángeles, Luca, Pablo, Silvia, Carmen y con los que he compartido numerosos congresos pero que han trabajado desde la distancia en el CIC-Biogune: Ana, Illaria y Dani.

También agradecer a Jim (The Scripps Research Institue, San Diego, CA, USA) y a Dusan (School of Chemistry, University of Edinburgh, Scotland, UK) por haberme abierto las puertas de sus respectivos laboratorios durante las estancias de investigación y haberme mostrado y compartido sus grupos multidisciplinares e internacionales de donde surgen los grandes proyectos científicos. Ha sido un placer haber estado por unos meses integrado en sus grupos y también haberme permitido conocer nuevas áreas científicas.

Durante estos años, el trabajo realizado no hubiera sido posible sin la colaboración con multitud de grupos de otros países y también del CIB a los que no tengo más que agradecer sus esfuerzos realizados y la posibilidad que me han dado de trabajar con ellos en proyectos conjuntos con resultados muy satisfactorios y que han enriquecido el trabajo realizado.

No sería justo si me olvidara de mis profesores de la carrera y del máster que me han educado en las bases de la ciencia química, bioquímica y farmacéutica en las que con entusiasmo he querido profundizar.

Por último, agradecer a todos mis familiares y amigos que me han apoyado en este período formativo, a veces duro, a veces llevadero pero que sin duda su apoyo ha sido necesario.

CONTENTS

Abbreviations

A

AHA: acetohydroxamic acid

AXHR: acute humoral xenograft rejection

B

BMRB: biological magnetic resonance data bank

C

catMMP12: catalytic domain of MMP-12

C/EBP: CCAAT/enhancer-binding protein

CHOP: C/EBP homologous protein

CLEANEX-PM: chemical exchange phase modulated experiment

CMP: cytidine mono-phosphate

CPZ: chlorpromazine

CRD: carbohydrate recognition domain

CSP: chemical shift perturbation

D

DCC: dynamic Combinatorial Chemistry

DCL: dynamic Combinatorial Library

DMSO: dimethylsulfoxide

DNA: deoxyribonucleic acid

DOSY: diffusion ordered spectroscopy

E

ECL: erythrina crystagalli lectin

ECM: extra cellular matrix

EPO-A: epothilone A

EPO-tc: trans-cyclopropane epothilone derivative

F

Fab: antigen binding fragment

FDA: Food and Drug Administration

Frq2: frequenin-2

G

Gal: galactose

gal-3: galectin 3

GalNAc: N-Acetylgalactosamine

GDP: guanidine di-phosphate

Glc: glucose

GlcNAc: N-Acetyl glucosamine

GTP: guanidine tri-phosphate

H

HA: hemagglutinin

HAR: hyper acute rejection

HB: hydrogen bond

hetNOE: heteronuclear NOE

HPA: helix pomatia agglutinin

HSQC: heteronuclear single quantum coherence

I

IDP: intrinsically disordered protein

IR: infra-red

ITC: isothermal calorimetry

K

K_D: dissociation constant

M

MALDI-TOF: matrix assisted laser desorption ionization – time of flight

MGL: macrophage galactose lectin

MMP-12: matrix metalloproteinase 12

MS: mass spectrum

MSA: microtubule stabilizing agents

MT: microtubule

MWCO: molecular weight cut off

N

NA: neuraminidase

N-CHOP: N-terminus domain of CHOP

Neu5Gc: 5-glycolyl neuroaminic acid

Neu5Ac: 5-acetyl neuroaminic acid

NMR: nuclear magnetic resonance

NNGH: N-isobutyl-N-(4-methoxyphenylsulfonyl) glycyl hydroxamic acid

NOE: nuclear Overhausser effect

P

PBS: phosphate buffered solution

PDB: protein data bank

PFG: pulse field gradient

Q

QCM: quartz crystal microbalance

R

R_1: longitudinal relaxation rate

R_2: transversal relaxation rate

ROESY: rotating frame NOE correlation spectroscopy

S

SPR: surface plasmon resonance

STD-NMR: saturation transfer difference NMR

T

T_1: longitudinal relaxation time

T_2: transversal relaxation time

TACA: tumor associated carbohydrate antigen

TBA: tubulin binding agents

τ_c: total correlation time

TF: transcription factor

TFA: trifluoroacteic acid

TLC: thin layer chromatography

TOCSY: total correlation spectroscopy

Tr-NOESY: transferred NOE spectroscopy

Trp: tryptophan

U

UDP: uridine di-phosphate

UV: ultra-violet

W

Water-LOGSY: water-ligand observed via gradient spectroscopy

WT: wild type

α

α-OMeGal: α-methyl galactopyranoside

α-OMeGalNAc: α-methyl N-acetylgalactosamine pyranoside

Resumen

El reconocimiento molecular está presente en numerosos procesos vitales y relevantes para la salud y la enfermedad. El conocimiento de la comunicación intermolecular a nivel estructural es de especial interés para el entendimiento de numerosos procesos biológicos. De entre la gran variedad de técnicas aplicadas al reconocimiento molecular, la Resonancia Magnética Nuclear (RMN) proporciona información valiosa a diferentes niveles de los bio-complejos formados.

Desde el comienzo, la RMN ha sido explotada hasta alcanzar límites insospechados. En esta tesis, diferentes técnicas, métodos y protocolos basados en RMN se han aplicado extensamente para estudiar una variedad de procesos de reconocimiento molecular de diferentes niveles de complejidad. En particular, esta aplicación de la RMN ha permitido recopilar información sobre las características del reconocimiento molecular de estos eventos abordado desde las perspectivas del ligando y/o receptor. Se ha hecho especial hincapié en desentrañar los fenómenos de interacción en los que intervienen los carbohidratos. De manera paralela, también se han utilizado métodos de RMN análogos en el campo de la dinámica de proteínas así como para cribar mezclas de ligandos mediante un enfoque de química dinámica combinatoria guiada por RMN.

Introducción

Este capítulo introduce las bases de la RMN aplicada al reconocimiento molecular y describe el fondo bioquímico de los sistemas biológicos estudiados. Las perspectivas de ligando y receptor basadas en RMN se presentan como alternativas para obtener acceso a información complementaria de la interacción ligando-receptor.

Desde el punto de vista del ligando, los experimentos de RMN están enfocados en la observación de las señales del ligando revelando información clave sobre el epitopo de unión del ligando así como de su conformación en el estado unido. Los experimentos de RMN más relevantes dentro de esta clase (STD-NMR) se basan en la transferencia de saturación de la magnetización para un sistema en intercambio químico (del estado libre al unido y viceversa). Por otro lado, los experimentos de RMN desde el punto de vista del receptor, proporcionan información a nivel atómico valiosa sobre receptor y permiten definir con mayor precisión rasgos clave del evento de reconocimiento (ej. amino ácidos del sitio de unión). Estos últimos métodos, generalmente implican la observación de las señales del receptor en experimentos de correlación heteronuclear en ausencia y en presencia del ligando.

También, se describen las principales interacciones intermoleculares esenciales para el reconocimiento molecular y se introducen los principales actores bioquímicos (proteínas, carbohidratos, ADN) involucrados en los procesos de reconocimiento.

Objetivos generales

El objetivo general de la tesis es la aplicación de métodos basados en RMN desde el punto vista del ligando y del receptor para el estudio de interacciones de ligandos de diversa naturaleza química con receptores de interés biológico y/o biomédico.

Aplicaciones de la RMN al reconocimiento molecular de carbohidratos

Este capítulo se encuentra enfocado a la obtención de información clave del reconocimiento de carbohidratos para aplicaciones biomédicas. Durante mucho tiempo, los carbohidratos fueron considerados como la principal fuente de energía de los seres vivos. Hoy en día, resulta evidente que los carbohidratos desempeñan un papel importante en numerosos procesos biológicos y que son capaces de transmitir ingente cantidad de información biológica importante. Los carbohidratos situados en la superficie de las células y su entorno se encuentran a menudo implicados en importantes procesos de reconocimiento molecular estableciendo interacciones con anticuerpos, proteínas, virus, células, bacterias… que desencadenar una variedad de procesos intra e inter celulares de interés fisiológico y patológico.

En esta sección, mediante la aplicación de estrategias basadas en RMN tanto del punto vista del ligando como del receptor, se ha estudiado diferentes sistemas biológicos carbohidrato-receptor: i) El análisis de las características de reconocimiento de los miméticos sintéticos de antígeno Tn reveló un modo de unión similar al modelo de referencia del Tn antígeno (α-OMeGalNAc) hacia diferentes lectinas; ii) Se ha demostrado el carácter bifuncional de un glicano ramificado sintético unido covalentemente a un inhibidor de proteasa y su capacidad para dirigir simultáneamente diferentes proteínas. iii) La especificidad de la interacción de anticuerpos monoclonales dirigidos al xenoantígeno basado en carbohidratos se ha deducido por STD-NMR, mostrando también la existencia de diferentes afinidades para el xenoantígeno α-gal y una serie de anticuerpos. En particular, también se ha mostrado la especificidad anomérica de la interacción del antígeno de ácido N-glicolilneuramínico (Neu5Gc) con diferentes anticuerpos. iv) El modo de unión de diferentes cepas de hemaglutininas (HA) del

virus de la gripe hacia los epítopos del ácido siálico presentes en los N-glicanos también ha sido elucidado por STD-NMR. Los datos experimentales, combinados con los cálculos de docking, han puesto de manifiesto la existencia de distintos modos de unión, dependiendo de la regioquímica del enlace Neu5Ac

Aplicaciones de RMN a otros sistemas y sistemas dinámicos

Los procesos biológicos son inherentemente dinámicos ya que ligandos y proteínas son capaces de ensamblarse en multitud de conformaciones. El entendimiento de este ambiente dinámico intrínseco es de suma importancia para monitorizar, explicar y modular los eventos de unión y reconocimiento molecular. En este contexto, los métodos de RMN proporcionan una diversidad de estrategias experimentales basadas en parámetros de relajación y / o interacciones dipolares para obtener información sobre el comportamiento conformacional de ligandos y receptores en sus estados libres y enlazados. Además, la RMN también puede aplicarse en sistemas químicos dinámicos, incluyendo la deconvolución de la interacción de bibliotecas combinatorias dinámicas con receptores, un proceso en el que se producen reacciones múltiples reversibles.

En este capítulo se describen diferentes ejemplos de aplicación de métodos de RMN al estudio de las características estructurales, conformacionales y dinámicas de sistemas de naturaleza química diversa e interés biológico. En particular, i) se han aplicado métodos de RMN para estudiar el grado de orden de dominios de proteínas mediante el análisis de los parámetros dinámicos de RMN de un fragmento del factor de transcripción CHOP. ii) Se ha realizado análisis conformacional por RMN basado en experimentos tr-NOESY de un derivado sintético de epotilona unido a tubulina revelando un modo de unión similar al descrito para la epotilona-A. iii) La interacción de una familia de moléculas con el surco menor del ADN se ha caracterizado mediante RMN mostrando su especificidad por regiones ricas en A / T. iv) Por último, la aplicación de métodos de RMN al estudio de una química combinatoria dinámica en presencia de una proteína determinada (Frequenin-2) ha permitido la identificación directa no ambigua de nuevos ligandos

Abstract

Molecular recognition is at heart of life processes relevant to health and disease. Deciphering the code of the intermolecular communication at structural level is enormous interest for the understanding of many biological events. Among the variety of techniques applied in this field, NMR may provide valuable information, at different levels, on the formed bio-complexes.

NMR has been continuously exploited to reach unexpected limits a few years ago. In this thesis, different NMR techniques, methods and protocols have been extensively applied to study a variety molecular recognition processes at diverse levels of complexity. In particular, these applications of NMR have permitted gathering information on the molecular recognition features of these events, from the ligand and/or receptor perspectives. Special emphasis has been paid to unravel interaction phenomena where carbohydrates are involved. In a parallel manner, analogous NMR methods have also been used in the field of protein dynamics as well as for screening ligand mixtures through a NMR-guided dynamic combinatorial chemistry approach.

Introduction

The initial section of this thesis outlines the basic concepts for applying NMR methods to study molecular recognition events and describes the biochemical background of the employed systems. The NMR-based ligand and receptor perspectives are presented as alternatives to gain access to complementary information of the ligand-receptor interaction.

The ligand-point-of-view NMR experiments, focused on the observation of the ligand NMR signals, reveal key information on the interacting epitope and ligand conformation. The most relevant NMR experiments within this class (STD-NMR) are based on the detection of simple saturation transfer phenomena for a system in chemical exchange (free to bound and viceversa). Conversely, the receptor point of view NMR experiments provide atomic level information on the receptor and permit to define more precisely key features of the recognition event (i.e. amino acids at the binding pocket). The NMR experiments from the receptor perspective, generally, involve the observation of receptor signals in heteronuclear NMR experiments before and after the addition of the ligand.

In addition, we also introduce the intermolecular interactions important for molecular recognition and the biochemical players (i.e. carbohydrates, proteins, DNA) that are involved in the interacting process.

General objective

The general objective of the thesis is to apply ligand- and receptor-based NMR methods to study the interaction of a variety of ligands of diverse chemical nature with receptors of biological and/or biomedical interest.

NMR applications to carbohydrate recognition

This part is focused on gathering key information of carbohydrate recognition relevant for biomedical applications. During long time, carbohydrates were mainly considered as the source of energy for living organisms. It is evident today that carbohydrates play additional important roles in numerous life processes and also transmit important biological information. Carbohydrates located at the surface and cellular surroundings are involved in molecular recognition processes by establishing interactions with antibodies, proteins, viruses, cells, bacteria… which trigger a variety of intracellular and extracellular processes of special physiological and pathological interest.

Herein, by applying NMR strategies that rely on ligand and/or receptor observation, we have studied different carbohydrate-based biological systems: i) The analysis of the recognition features of synthetic tryciclic Tn antigen mimetics revealed a similar binding mode as that of the Tn reference model (α-OMeGalNAc) towards different lectins. ii) The bifunctional character of a synthetic branched glycan covalently connected to a protease inhibitor and its capacity for targeting simultaneously different proteins has been demonstrated. iii) The specificity of the interaction of monoclonal antibodies targeting the carbohydrate-based xenoantigen has been deduced by STD-NMR, also showing the existence of different affinities for the α-gal xenoantigen and derivatives thereof to a series of antibodies. In particular, the anomeric specificity of the interaction of the N-glycolylneuraminic acid (Neu5Gc) antigen towards different antibodies has also been shown. iv) The binding mode of the hemagglutinins (HA) from influenza virus strains towards the sialic acid epitopes present in N-glycans has also been elucidated by STD-NMR. The experimental data, combined with docking calculations,

have revealed the existence of distinct binding modes, depending on the regiochemistry of the Neu5Ac linkage.

NMR applications to dynamic processes and other systems

Biological processes are inherently dynamic since proteins and ligands can sample a vast ensemble of conformations. The understanding of the intrinsic dynamic features of the system under study is of paramount importance to monitor, explain and modulate the associated binding events. In this context, NMR provides a diversity of experimental strategies based on relaxation parameters and/or dipolar interactions to get insights into the conformational behavior of ligands and receptors in their free and bound states. Moreover, NMR can also be applied in chemical dynamic systems, including the deconvolution of the interaction of dynamic combinatorial libraries with receptors, a process where multiple reversible reactions take place.

In this chapter, we describe different examples of the application of NMR methods to the study of the structural, conformational and dynamics features of systems of diverse chemical nature and biological interest. In particular, i) NMR methods have been applied to study the degree of ordering of protein domains by the analysis of NMR dynamic parameters (R_1, R_2 and hetNOE) of a fragment of the CHOP transcription factor. ii) In a parallel manner, the NMR-based conformational analysis (tr-NOESY experiments) of a synthetic epothilone derivative bound to tubulin revealed the presence of a similar binding mode to that described for natural epothilone-A. iii) The interaction of a family of molecules with the DNA minor groove has been also characterized by other NMR methods, allowing to deduce their specificity towards the A/T rich regions as well as the structure of their complexes. iv) Finally, the application of NMR methods to the study of a dynamic combinatorial chemistry in the presence of a given protein target (Frequenin-2) has permitted the direct non-ambiguous identification of new binders.

Chapter 1

Introduction

1.1. Molecular recognition and NMR

Molecular recognition is at heart of life processes relevant in health and diseases. The deciphering of the code of the intermolecular communication at structural level has been revealed of enormous interest for the understanding of many biological processes such as biochemical cascades, enzymatic reactions, metabolism, fertilization, immune reactions, pathogen infection, and tumor metastasis among others. Over the years numerous techniques (isothermal calorimetry (ITC),[1–3] quartz crystal microbalance (QCM),[4,5] surface Plasmon resonance (SPR),[6,7] X-ray crystallography,[8,9] nuclear magnetic resonance (NMR)[10–14]…) have been developed for the study of these interactions.

From a pharmaceutical point of view, the deep knowledge of the key interactions is of special importance for the rational design of proper drugs for therapeutic targets. Thus, the study of the molecular recognition has experienced a large development. The future perspective of molecular recognition is promising given the fast development of the technologies and the new information that will continuously be generated. As example of that, the supramolecular chemistry field has matured in the last years passing from the fundamental molecular recognition research that studied the formation of supramolecular complexes by non-covalent forces between species to the generation of molecular machines, awarded with the Nobel Prize in Chemistry in 2016.

The structure of biomolecules and its dynamic behavior, in the context of the cellular medium, have a key role in molecular recognition events and hence in the biological activity. Therefore, the study of molecular recognition in solution is of central importance.

Among the variety of techniques applied in molecular recognition studies, NMR provides valuable information at atomic level of the formed bio-complexes upon molecular recognition events in solution. This approach nicely complements the constraint environments required for other alternatives as X-ray crystallography.

1.1.1. NMR in molecular recognition

Since the first observation of the nuclear magnetic resonance (NMR) in the mid-1940s by Purcell and Bloch (awarded with Nobel prize in 1952), NMR has demonstrated to be a disruptive methodology in science endorsed by three additional Nobel prizes (Richard E. Ernst, 1991; Kurt Wüthrich, 2002 and Paul C. Lauterbur and Peter Mansfield, 2003). NMR spectroscopy was firstly employed in organic chemistry for structural elucidation of synthetic

compounds, but numerous additional applications have been developed concerning other fields such as biochemistry[15–17] or inorganic chemistry.[18,19] However, over the years NMR (figure 1) has experienced a significant revolution with three major achievements because of technology improvement: 1. The early introduction of the Fourier transformation that decrease the acquisition times compare with the previous continuous wave method, 2. The development of multidimensional NMR, and 3. The introduction of cryo-technology that increase significantly the magnetic fields. The generalization of other breaking methods, such as Dynamic Nuclear Polarization is still under development.

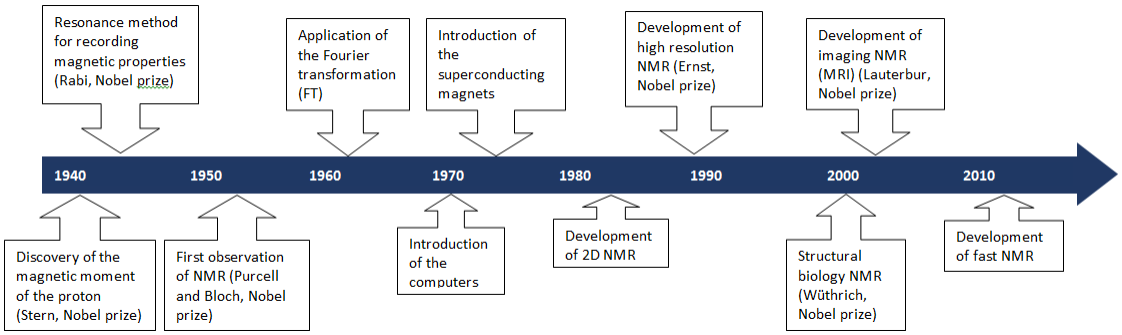


Figure 1. NMR timeline of the last 70 years. NMR hits are annotated.

NMR[20,21] is based in the intrinsic magnetic properties of several atomic nuclei that possess non-zero nuclear spin quantum number, **I**. The property of nuclear spin is necessary for the resonance phenomenon. Therefore, the spinning nuclei possess angular momentum, **P**, with a concomitant magnetic momentum, $\boldsymbol{\mu}$. **P** is directly related with $\boldsymbol{\mu}$ by the magnetogyric ratio, γ, which is characteristic of each atomic nucleus (equation 1).

$$\mu = \gamma P$$

Equation 1. Relationship between the angular momentum and the magnetic momentum of a nuclear spin.

According with the quantum rules, the magnetic momentum and the angular momentum are quantified. Thus, for a spin quantum number **I**, exist 2**I**+1 possible spin sub-states comprising from -**I** to +**I** values, for instance for a spin **I**=½ such as ^{1}H, ^{13}C or ^{15}N, there are two possible degenerated energy levels ½ and –½ dubbed as α and β. In the presence of an external magnetic field, B_0, the degeneracy disappears splitting into two different energy levels. The energy difference between the two states is directly proportional to the strength of the magnetic field, B_0, and is given by equation 2.

$$\Delta E = \gamma \hbar B_0$$

Equation 2. Energy between spin levels when applied a magnetic field, B_0.

The occupancy of the α and β energy levels is driven by the Boltzmann distribution (equation 3).

$$\frac{N_{low}}{N_{upp}} = e^{\frac{-\Delta E}{kT}}$$

Equation 3. Boltzmann distribution of the populations of the spin levels.

The asymmetric distribution of the nuclear spin in the two states, α and β, in the presence of an external magnetic field is the basis of the observation of NMR signals. When the radiation frequency matches with the energy gap between levels, it stimulates a transition from the fundamental state to the state of higher energy taking place the resonance condition. Noteworthy, the Boltzmann distribution is temperature dependent and the ΔE between spin states is rather low even at high magnetic fields. Thus, the populations difference are similarly small being NMR very insensitive relative to other spectroscopies such as IR o UV. Every energy quantum is described by $\Delta E = h\nu$, which in NMR is transformed into $\omega_0 = \gamma B_0$ where $\boldsymbol{\omega}$ is the rotation frequency that the nucleus precesses (figure 2) with a circular path about the applied magnetic field, Larmor frequency.

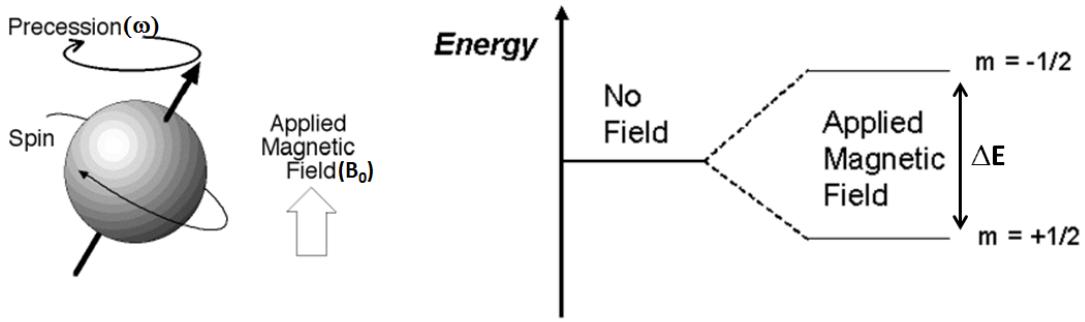


Figure 2. Diagram of the energy levels of a nucleus of spin I=1/2 when an external magnetic field is applied.

Therefore, the Larmor frequency of each nucleus depends directly on the magnetogyric ratio and the strength of the magnetic field. However, in a molecule, the effective magnetic field, B_{eff}, experienced by the nuclei of different atoms is affected by the nearby electrons of the chemical bonds and the chemical environment due to the shielding or de-shielding.

Consequently, the different nuclei of a molecule precess with a particular Larmor frequency depending on their chemical environment (equation 4).

$$\omega = \gamma B_{eff}$$

Equation 4. Larmour frequency (ω) of an atomic nucleus in the presence of a magnetic field, B_{eff}.

Hence, the Larmor frequencies are translated into terms of chemical shifts, δ, by equation 5.

$$\delta\ (ppm) = \left(\frac{\omega - \omega_{ref}}{\omega_{ref}}\right) 10^6$$

Equation 5. Definition of chemical shift.

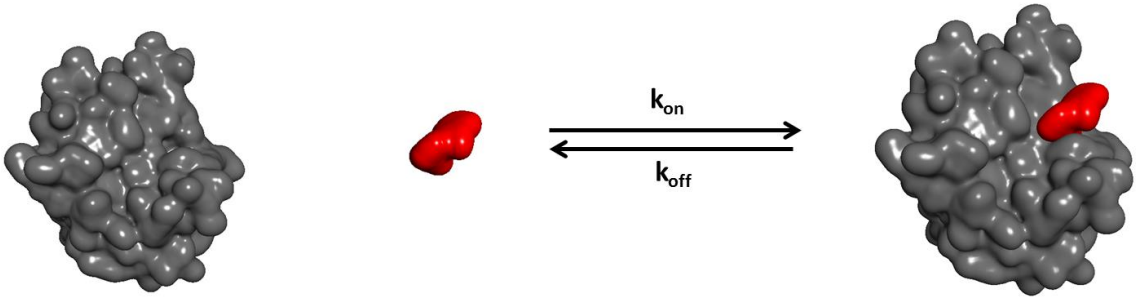
Where ω is the Larmor frequency of the nucleus and ω_{ref} is the reference Larmor frequency of a reference compound.

From the beginning, NMR has been exploited to unexpected limits being in numerous cases preferred to other spectroscopic techniques even though its insensitivity. The great development of the NMR in the last decades has allowed the transition of the NMR from the structural organic chemistry to the biochemistry fields arising as a powerful and versatile technique for the study of biochemical processes thanks to the development of an immense variety of experiments able to extract detailed information at atomic level.

1.1.2. Molecular recognition: Chemical equilibrium and kinetics exchange[22]

Binding events are defined according to thermodynamics as an equilibrium between the free species and the bound complex. For a 1:1 ligand:receptor complex, the equilibrium is given by equation 6.

$$[R]+[L] \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} [RL]$$



Equation 6. Equilibrium at the binding event.

Where **[R]** is the receptor concentration, **[L]** the ligand concentration, $\mathbf{k_{on}}$ is the rate constant to form the RL complex while $\mathbf{k_{off}}$ is the rate constant for complex dissociation, which is inversely proportional to the lifetime of the bound complex. Therefore, the binding affinity is described by the dissociation constant ($\mathbf{K_D}$) as:

$$K_D = \frac{[R]\,[L]}{[RL]} = \frac{k_{off}}{k_{on}}$$

Equation 7. Expression of the dissociation constant (K_D).

Thus, a strong binder (high affinity) with a long complex lifetime and concomitant low k_{off} has a small K_D whereas weak binder (low affinity) displays larger K_D. As a consequence, strong ligands with small K_D values saturate receptor sites at lower concentrations than weaker ligands. Under these conditions, any observed NMR parameter tends towards its endpoint, RL complex formation, very rapidly. Conversely, weak ligands have to be present in significant excess over the receptor to achieve significant populations of the bound complex and, hence, to induce an observable change on the NMR parameter. Experimentally, these considerations relative to the different K_D limits are important for detection of ligands with different affinity.

Most of the NMR observables used for the molecular recognition involved a change of NMR parameters as a result of receptor-ligand association. However, the kinetics of the binding and the residence time of the ligand at the binding pocket usually determine de final observable value of the NMR parameter. Consequently, if the observed NMR species are in rapid on-off exchange, the observed NMR parameter value may be, then, described by the molar fraction-weighted contributions of the different species, free and bound, present in the binding equilibrium. For a two states system, equation 8 can be applied.

$$M_{observed} = \chi_{free} M_{free} + \chi_{bound} M_{bound}$$

Equation 8. Observed NMR value in a rapid exchange equilibrium for a two-states system.

Where **M** is the observable NMR parameter and **χ** is molar fraction. Thus, the bound population must be sufficient to induce a variation on the NMR observable parameter. This situation is very important for the case of chemical shift; in the fast-exchange regime, one unique signal will be observed, located at the molar fraction-weighted average between both end states (equation 8).

In the case of slow on-off exchange, the chemical shift will display different values for each possible form. In this case, their signal intensities will be proportional to the corresponding molar fractions (equation 9).

$$M_{observed\,1} = \chi_{free} M_{free}$$

$$M_{observed\,2} = \chi_{bound} M_{bound}$$

Equation 9. Observed NMR values in a slow exchange equilibrium.

However, in the intermediate on-off exchange, a new component related to the exchange kinetics contributes to the final shape and intensity of the NMR observed parameter. A complex function results from the contributions of the bound and free forms and their exchange kinetics (equation 10).

$$M_{observed} = f\left(\chi_{free} M_{free}; \chi_{bound} M_{bound}; k_{exchange}\right)$$

Equation 10. Observed NMR values in an intermediate exchange equilibrium.

1.1.3. NMR parameters in the context of intermolecular complex, equilibrium, dynamics, structure and size

Different NMR observables are useful to monitor binding events. As already mentioned, the value of the NMR observable depends on the equilibrium exchange rate. Below, we described some of the most common NMR observables used for analyzing molecular recognition processes.

1.1.3.1. Chemical Shift

The chemical shift is very sensitive to the chemical environment of the observed nuclei. Thus, if a ligand binds, its own chemical environment as well as that of the receptor binding pocket, changes. This fact induces perturbations in the chemical shift of the nuclei of both binding partners. Indeed, the observation of changes in the chemical shift in one binding partner upon

addition of the other may be observed in either monodimensional spectra or 2D NMR spectra. Due to the different on-off exchange rates in ligand binding, different chemical shift perturbation effects (figure 3) may be described, depending on the presence of fast, slow or intermediate exchange (equations 8-10).

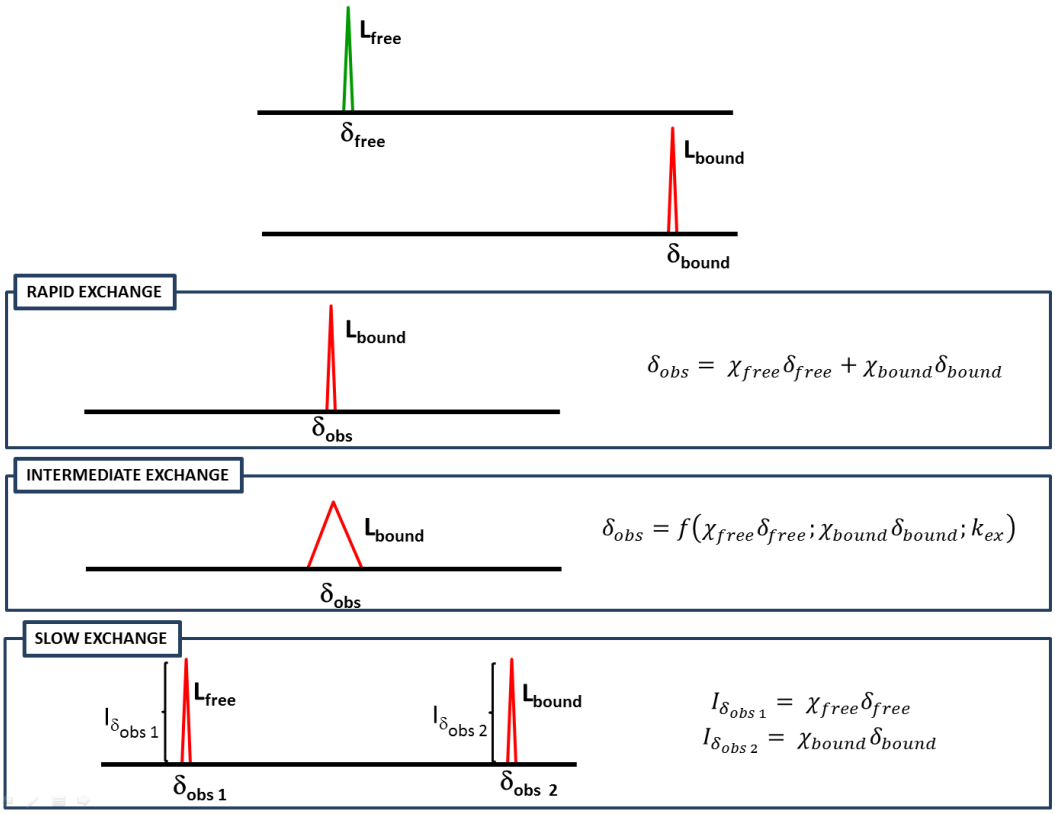


Figure 3. Scheme of the chemical shift variation depending on the exchange rate.

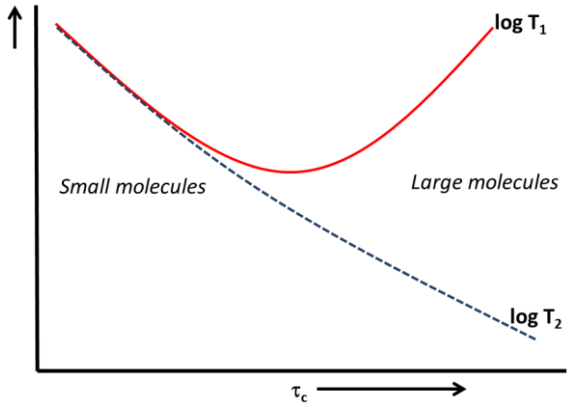
However, changes in the chemical shifts may be also caused by changes in the solvent environment (pH, ionic strength of the buffer or temperature) or intrinsic ligand and/or receptor conformation. Thus, a precise sample handling helps in the discrimination of the chemical shift perturbations of relevance for the binding event over those due to changes in the solvent environment.

1.1.3.2. Relaxation rates

After any perturbation of an external force, the perturbed system takes time to return to its equilibrium state. In NMR experiments, two main time-dependent magnitudes, relaxation modes, are of interest depending on the orientation where the magnetization of the observed nuclei is located with respect to the static magnetic field of the spectrometer. Thus, longitudinal relaxation ($R_1=1/T_1$) and transverse relaxation ($R_2=1/T_2$) rates are described

corresponding to the evolution of the magnetization when parallel to the external field (z-axis) or orthogonal to it (x,y plane) respectively.

The relaxation rates of nuclear spins depend on the global rotational correlation time (τ_c) of the molecule, directly related with the molecular size in solution, and hence on the molecular weight. Thus, ligands with low molecular weight display large T_1 (small R_1) and large T_2 (small R_2) with a concomitant slow longitudinal and transverse relaxation rates. In contrast, receptors with large molecular weight display large T_1 (small R_1) and small T_2 (large R_2) with a concomitant slow longitudinal relaxation and fast transverse relaxation rates (figure 4).



Equation 11. Representation of the relaxation times evolution with τ_c.

In the molecular recognition event, the bound ligand adopts the relaxation properties of the receptor and transiently retains these properties after dissociation, depending on its relaxation rates. The observation of the change of the ligand NMR properties is often used as indicative of binding.

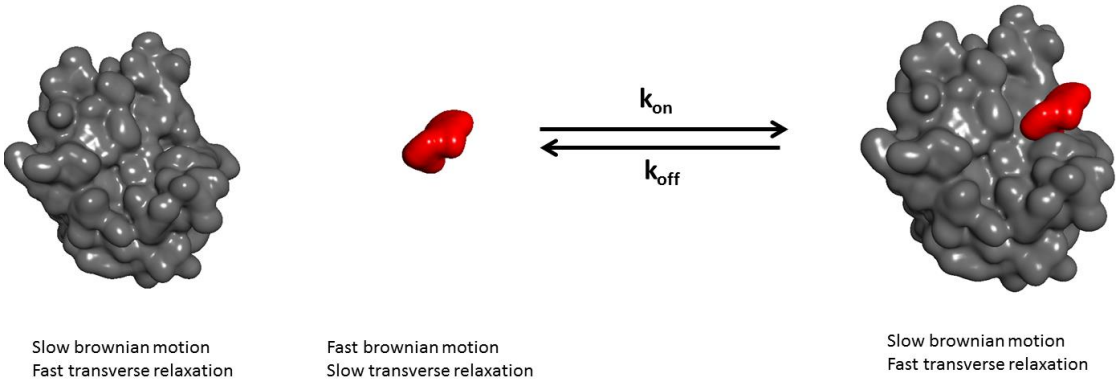


Figure 4. Relaxation properties of the species in the binding equilibrium.

1.1.3.3. Dipolar interactions: Nuclear Overhauser Effect (NOE)

The nuclear Overhauser effect (NOE) has been largely exploited in NMR for conformational analysis studies. The NOE is based in the dipolar coupling of the nuclear spins through the space. Dipolar coupling or dipole-dipole interaction occurs via cross-relaxation, a process through which an excited nuclear spin reverts to its equilibrium populations affecting other closer spins by dipole-dipole interaction. Since cross-relaxation has a strong dependency with the distance, the NOE intensity is actually proportional to r^{-6}. Thus, the inter-nuclei distance is critical for the observation of NOE, which provides information on the 3D structure and conformation. Furthermore, the NOE intensity sign depends on the global rotational correlation time (τ_c) of the molecule. Thus, small molecules with a short τ_c display positive NOE, while large molecules with large τ_c have opposite sign (figure 5).

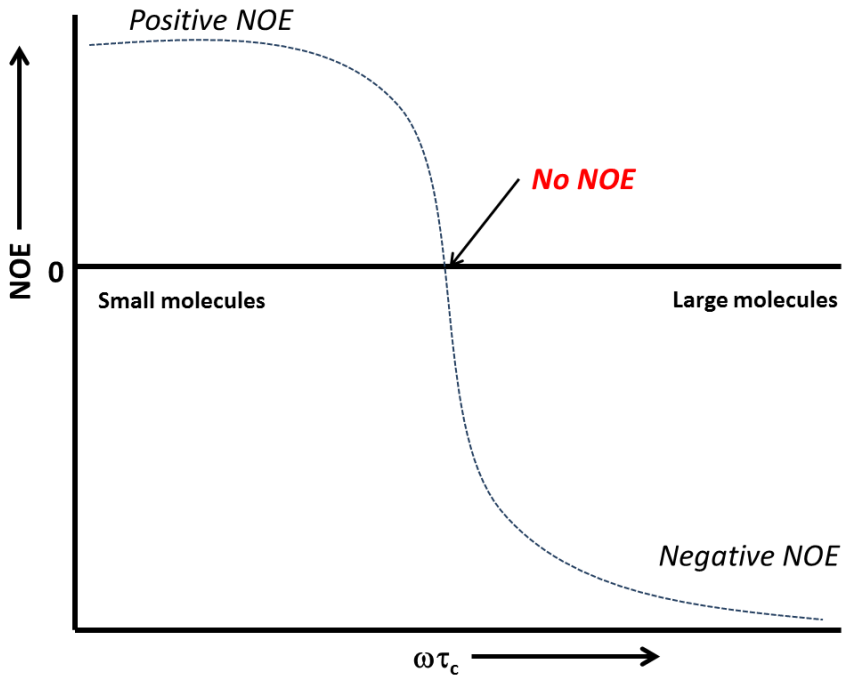


Figure 5. NOE dependency with the total correlation time (τ_c) and spectrometer frequency (ω).

Based on the characteristics of the NOE, in addition of its extensive use in conformational analysis, it has been used for the identification of molecular recognition events. For instance, the transferred NOESY experiment (tr-NOESY) applies the simple observation of sign change in ligand NOESY cross peaks (figure 6).

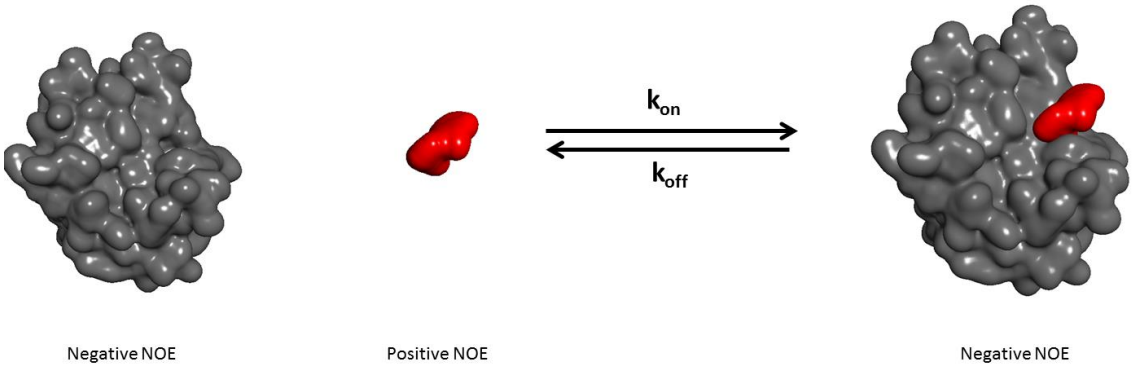


Figure 6. NOEs of the interacting species in the binding equilibrium.

1.1.3.4. Hydrodynamic diffusion coefficient: Diffusion ordered spectroscopy (DOSY)[23]

With the development of reliable gradient technology for NMR the investigations of diffusion processes by pulsed-field gradient (PFG) became possible. Diffusion can be either investigated by analysis of the relaxation data or by PFG.[24] However, the PFG based experiments are sensitive to motions at the ms to s time-scale, matching with the molecular recognition time-scale. Diffusion ordered spectroscopy (DOSY) is a PFG based experiment that facilitates the study of mixtures by the different diffusion properties of the components. DOSY has been further applied to study binding of low-molecular weight ligands to large receptors.[25–28] The main concern is that the ligand has to reside significant time in the binding pocket to detect significant changes in the diffusion coefficient since the observed diffusion coefficient consists in the linear average of the diffusion constant of the ligand in the free and bound state (equation 6).

$$D_{observed} = \chi_{free} D_{free} + \chi_{bound} D_{bound}$$

Equation 12. Diffusion coefficient observed of a binder in DOSY experiments.

Where **D** is the translational diffusion coefficient.

The diffusion coefficient is defined by the Stokes-Einstein equation being inversely proportional to the hydrodynamic radius of the molecule (equation 13).

$$D \propto \frac{1}{r}$$

Equation 13. Relation between the hydrodynamic radius (r) and the diffusion coefficient (D).

Therefore, recognition by the receptor strongly affects the hydrodynamic behavior of the ligand inducing changes on its diffusion coefficient. This fact is observable by the decrease of the diffusion coefficient in the DOSY experiment. Obviously, receptor diffusion coefficient is also affected but the much smaller diffusion coefficient of the receptor, D_{free}, makes the contribution of the bound ligand, D_{bound}, insignificant.

1.1.3.5. Different time-scale processes observable by NMR

One of the advantages that have favored NMR to establish as a versatile tool to study dynamic systems is its large time-scale window. The NMR time scale rangs from the ps to s. Many key biochemical processes such as protein motions, ligand binding, or enzyme kinetics occur at these times. Thus, different NMR parameters provide essential information for these studies.

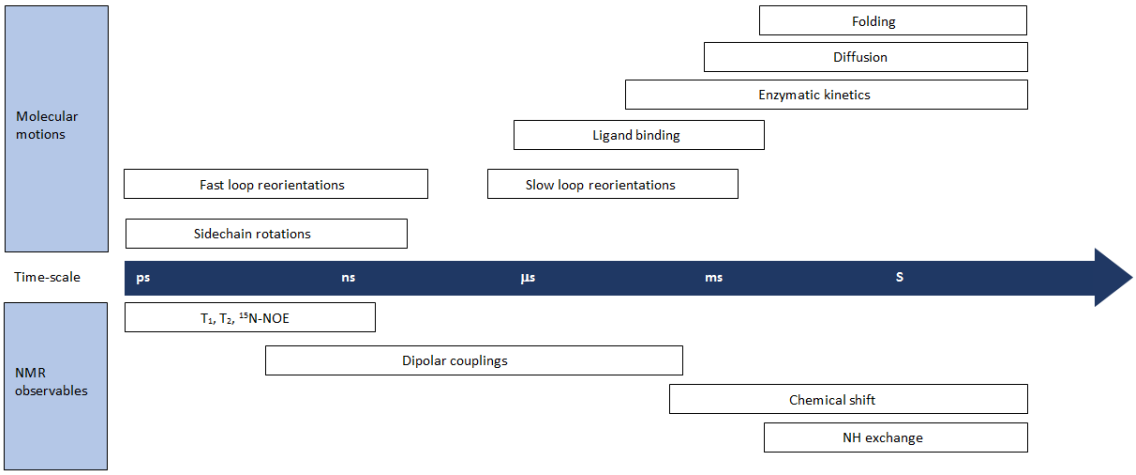


Figure 7. NMR parameters and molecular motions at the NMR time scale.

1.1.4. NMR experiments in molecular recognition

The study of molecular recognition events can be addressed from different perspectives. A variety of NMR experiments can be applied for gathering the required information. Two perspectives are commonly employed to describe the recognition phenomenon between partners. The first one involves the observation of the ligand properties. The second one, the observation of the receptor properties. Both perspectives are widely used and provide complementary information of the molecular recognition event.

1.1.4.1. The ligand point of view[29,30]

The ligand NMR perspective is often used for the study of weakly binding ligands. Normally, ligands are small molecular weight molecules whose characteristic NMR properties change with the association to large molecular weight macromolecules, proteins or DNA fragments. As already mentioned, small molecules, due to their size, display a fast Brownian motion, associated with a rapid tumbling rate, a concomitant fast rotational motion correlation time, and slow relaxation. Opposite properties, slow Brownian motion, slow tumbling, and large rotational motion correlation times, associated with fast transverse relaxation, are characteristic of macromolecules. Therefore, binding of ligands to macromolecules may be detected by the observation of increased transverse relaxation rates in the presence of the receptor. Qualitatively, these relaxation changes are easily observed by monitoring the line width of ligand signals in the free-state and in the presence of the receptor. In fact, the observed linewidth is inversely proportional to T_2. Thus, the observation of an increase of the line width in a simple ^{1}H NMR 1D spectrum is the first indication of binding (figure 8).

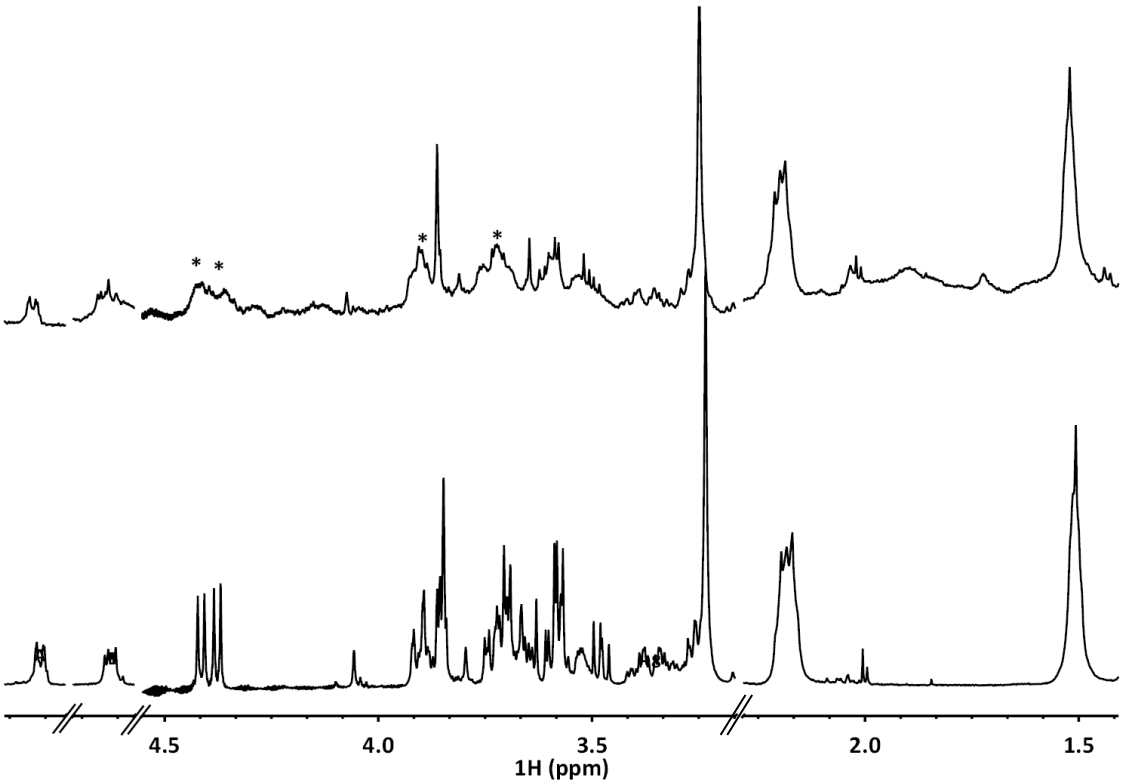


Figure 8. Example of line broadening as diagnostic of binding. On the top: ^{1}H spectrum of a ligand (Mw=1750 Da) in the presence of its receptor (Mw=30 KDa). On the bottom: ^{1}H spectrum of the ligand in absence of receptor.

Furthermore, changes in the chemical shifts may be observed in the simple ^{1}H NMR spectrum, which are also indicative of binding.

1.1.4.1.1. Saturation transfer experiments

Additionally, other methods based in a selective transfer of magnetization during the ligand residence time at the receptor binding pocket are frequently employed in ligand-based NMR methods. The saturation transfer experiments take advantage of the selective transfer of magnetization from one of the partners in the binding event (the receptor in this case) to the other partner (the ligand). This transfer perturbs its populations of the α and β states and can be easily detected in the NMR signals.

1.1.4.1.1.1. Saturation transfer difference NMR (STD-NMR)

The STD-NMR experiment is a robust and powerful experiment for identifying ligand-receptor interactions. It has been widely used for molecular recognition and it is suitable for NMR screening of ligand libraries. Indeed, it is of special relevance in the pharmaceutic industry for screening and fragment based drug discovery (FBDD),[31–34] since STD-NMR allows from one side, the direct observation of the bound ligand filtering the NMR signals for those non-binders. Moreover, it provides epitope information.

STD-NMR[35,36] is based on the transfer of magnetization between the receptor, which is selectively irradiated by a train of radiofrequency pulses, and any ligands bound to it. The STD spectrum, as its name indicates, is a difference of two 1D ^{1}H NMR spectra: the on-resonance spectrum, in which the saturation pulse is applied selectively on the receptor signals, and the off-resonance spectrum, in which this pulse is applied away from any receptor and ligand signals. Typically, in the case of protein receptors, the on-resonance irradiation pulse is set around the aliphatic region (from −1 to 1 ppm) or alternatively around the aromatic region (usually 6–8 ppm) if the ligand has no aromatic signals.[28] The saturation quickly propagates along the macromolecule via spin diffusion (dipolar cross relaxation mechanism), and then from the macromolecule to the ligand's nuclei in close contact with the macromolecule ($r < 4–5$ Å). The dissociation of the ligand, according to binding exchange kinetics, allows saturated ligand molecules to accumulate, and when the NMR spectrum is recorded, the saturated nuclei give rise to attenuated signals due to the inversion of the spin populations induced by the transferred saturation (figure 9). As blank, off-resonance spectrum is acquired. In the off-

resonance spectrum, the saturation pulse is set at a frequency where no signals of either receptor or ligands appear (e.g. 100 ppm). The subtraction of the on-resonance from the off-resonance spectra gives a difference spectrum showing the signals of those nuclei of the ligand that had received magnetization during their binding to the macromolecule. Analysis of these signals not only permits the identification of binding ligands among a mixture of compounds challenging the receptor, but also provides information of their binding epitope (figure 10), since nuclei of complex ligands, which are closer to the protein, become more saturated than those that do not participate directly in the binding.[37]

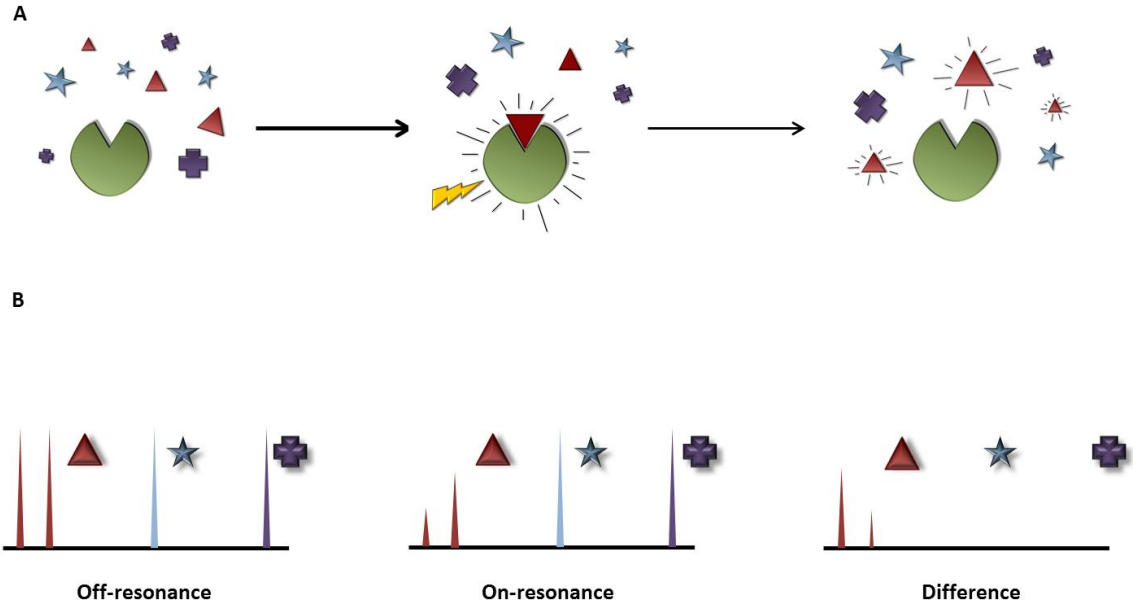


Figure 9. Scheme of the STD experiment for binder identification. Panel A: Selective transfer of magnetization from the irradiated protein to a bound ligand (red triangles). Panel B: Simulated STD spectrum, in which saturation of the bound molecule leads to a decrease of its NMR signals in the on-resonance spectrum. The STD-difference spectrum reveals only the NMR signals of saturated molecules.

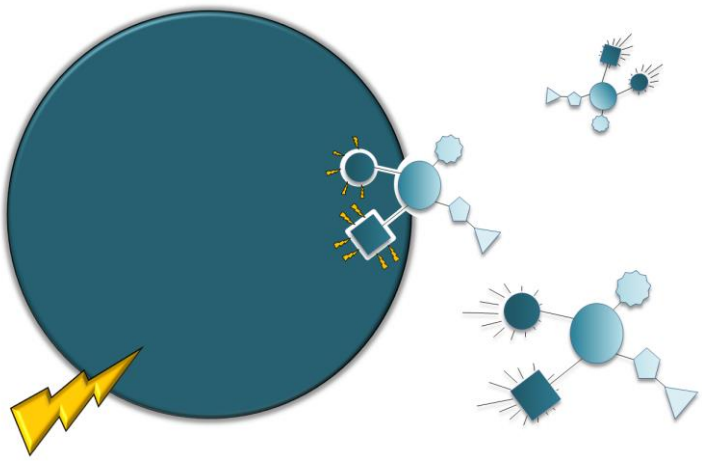


Figure 10. Scheme of the STD information for epitope mapping. Closer nuclei receive higher saturation from the receptor while nuclei far from the receptor are not affected.

Typically, STD-NMR experiments require small concentrations of receptor (low micromolar) in the presence of a large excess of ligand (50 to 100 fold excess), making this technique very convenient in terms of protein economy. From the information given, bound ligands may be efficiently detected and epitope mapping inferred and, in favorable cases, the binding affinities can be quantitatively estimated from initial saturation rates measured by ligand titration at different saturation times.[38,39] However, STD-NMR presents limitations. Due to its intrinsic dependency on fast exchange, STD-NMR is only useful for systems with K_D in the 10^{-8} to 10^{-3} M range, thus hampering the direct observation of high affinity ligands. Nevertheless, these ligands may be detected by means of competitions experiments with weaker and observable ligands. In addition, STD-NMR based competition studies with a well-known ligand not only permits the elucidation of the binding site, in the case of a competitive inhibition, but also the estimation of apparent binding affinities of other ligands by competitive titrations.

1.1.4.1.1.2. Water-LOGSY

In the Water-LOGSY NMR experiment the saturated entities are the bulk water molecules.[40,41] Water molecules at the surface of the receptor's binding pocket display similar characteristics to those of large molecules, whereas solvation water molecules of ligands behave like small molecules. Since the nuclear Overhauser effect (NOE) is related to the global rotational correlation time of the system, solvating water molecules adopt similar NOE properties than the solvated molecule, receptor or ligand, and display negative or positive intermolecular NOEs, respectively.

The WaterLOGSY experiment takes profit of the transfer of magnetization of bulk water molecules via different pathways. One pathway involves ^1H–^1H cross-relaxation mechanisms between the bound ligand and the water molecules of the binding pocket. A second pathway occurs via cross-relaxation with exchangeable protons of the receptor (i.e. of amino (NH) or hydroxyl (OH) groups), which propagate inverted magnetization to bound ligand protons.[42] These pathways of magnetization transfer act constructively. In all cases, inverted magnetization is transferred during the mixing time from the water molecules at the binding pocket to the bound ligands, yielding a change of sign of their NOEs to negative, characteristic of large molecules. In contrast, non-binders maintain the positive sign of their NOEs. The resulting spectrum displays signals of one sign for bound ligands that have received saturation during the mixing time and signals of the alternative one for non-binders.

The experimental requirements are similar to those of STD, i.e. small amounts of protein and excesses of ligand (up to 50-fold excess). However, the samples must be prepared in non-deuterated water. Information obtained through WaterLOGSY experiments is comparable to that obtained through STD-NMR. In contrast with STD-NMR, WaterLOGSY is less dependent on spin diffusion, being the preferred method when studying receptors of low proton densities, such as nucleic acids.[41]

1.1.4.1.2. Transferred NOESY (tr-NOESY)

Since its introduction by Bothner-By in the early seventies, tr-NOE NMR spectroscopy[43,44] has been thoroughly used to study receptor–ligand interactions. This technique focuses on the change of sign in the ligand NOE peaks upon binding to the receptor. This is due to the dependence of the NOE sign on the global motion rotational correlation time of the molecule, which in turn, is related to its molecular size. As in any NOE-based experiment, transferred NOEs provide information on the conformation of the ligands, now in their bound states. The tr-NOESY experiment is a regular 2D NOESY experiment in which the cross-peaks of the ligand change their NOE sign by virtue of the fact that, in the bound state, they adopt the motional properties of the macromolecule. As a result, non-binders show no NOE sign change upon the presence of receptor and their NOEs remain positive, while binders show negative NOE cross-peaks. Of note, the observed NOE is a contribution of both the free (positive NOE) and the bound ligand (negative NOE), but the cross-relaxation rate of the bound ligand is much larger than that in the free-state. Thus, the observed NOEs show a predominance of the contribution arisen by the bound form, even for low populations of the complex. Thus, the tr-NOESY experiment not only helps to detect interactions, but also provides geometrical information of the conformation of the ligand in the bound state.

1.1.4.1.3. Other Nuclei

Biochemically, few nuclei are of interest apart from proton, oxygen, carbon or nitrogen which are present in almost all biomolecules. Unfortunately, only ^{1}H is the natural abundant isotope with nuclear spin **I**= ½ and high magnetogyric ratio. However, carbon and nitrogen have NMR active isotopes with **I**=½, ^{13}C and ^{15}N respectively, even though their abundance is immensely low (1% for ^{13}C and less than 0.5% for ^{15}N). Nevertheless, they are useful observables for ^{15}N-labeled protein NMR and direct ^{13}C NMR detection methods.

However, other minor nuclei of particular interest have been explored such as ^{31}P or ^{19}F, both NMR active with a nuclear spin of **I**= ½ and natural abundant. Nevertheless, ^{31}P-NMR is rather limited to clinical applications i.e. to assess metabolites.[45] On the other hand, ^{19}F has been exploited extensively in medicinal chemistry and is present in numerous bioactive drugs and carbohydrates mimetics.[46–49] Thus, its importance in the pharmaceutic industry has enhanced ^{19}F NMR development.[50,51]

^{19}F NMR displays similar sensitivity to that observed for ^{1}H due to their similar magnetogyric ratios. The large spectral width allows the simultaneous screening of complex mixtures of molecules with minor signal overlapping. Additionally, the large chemical shift anisotropy of ^{19}F makes it very sensitive to binding to a receptor. From the ligand perspective, ^{19}F show additional advantages to those already mentioned, simplifying the interpretation of the spectra, since biomolecules does not incorporate fluorine. Thus, only signals of the ligand are observed without the introduction of any filter to remove the undesirable receptor signals. ^{19}F relaxation based experiments[52] and ^{19}F STD-NMR experiments[53] have also been introduced for studying ligand binding.

1.1.4.2. The receptor point of view

As mentioned before, although the ligand NMR point of view offers a variety of procedures for understanding the ligand behavior in the bound state, no direct information of the receptor may be extracted. In addition, the adoption of a ligand point of view NMR approach is desirable for weakly bound ligands. The gained information fades away for stronger affinities. Thus, the molecular recognition process can also be addressed from the NMR receptor perspective. It is well established and has been applied to proteins and large receptors since the early 1990s. Normally, the receptor point of view methodologies involve 2D NMR with isotopically ^{15}N and/or ^{13}C labeling of the receptor due to the complexity of handling the huge number of resonances in the ^{1}H NMR spectra. Therefore, it can be used for gathering information on conformational changes, binding pocket residues, affinity constants or other related receptor features.

1.1.4.2.1. Chemical Shift Perturbations. ^{1}H-^{15}N HSQC experiments

The current advances in protein expression have permitted the development of ^{1}H-^{15}N heteronuclear spectroscopy to assess information of the backbone amino acids changes upon

ligand binding. The chemical shift perturbation (CSP) analysis has emerged as the most frequently used method to study molecular recognition events from the receptor perspective. However, different challenges related with protein expression and the assignment of the amino acid resonances, which is very time-consuming and expensive, should be previously overcome. Protein labeling strategies and protein NMR have allowed the detailed study of domains up to 40 KDa.[28,54] However, partially labeled domains of larger proteins are also feasible and enough to extract the information of interest even though the limits are pushing further.[30]

The monitoring of the CSP of the protein resonances in a ^{1}H-^{15}N HSQC spectrum, as a result of a change in the chemical environment of the amino acids induced by ligand binding, provides information of the amino acids involved in the molecular recognition event. This methodology involves the acquisition of two protein ^{1}H-^{15}N HSQC spectra, before and after ligand addition. The simple monitoring of the resonances shows up the amino acid comprising the binding pocket but also K_D could be extracted by titration with the ligand. The observed CSPs could be quantified as an average of the perturbed nuclei (^{1}H and ^{15}N) with a correction factor due to the different magnetogyric ratios.

In comparison with the already ligand-observed described procedures, protein ^{1}H-^{15}N HSQC CSPs provide information of the receptor binding pocket but in contrast it is more time- and material-consuming and, due to the requirement of isotopically labeled proteins, it is also more expensive. Experimentally, this procedure is applicable to proteins of moderate size (up to 40 KDa) not only because of ^{15}N labeling limits but also due to unfavorable relaxation properties of larger macromolecules and its concomitant signal linewidth. Nevertheless, it allows the study of a wider range of ligand affinities.

1.1.4.2.2. Chemical Shift Perturbations. NOESY/TOCSY experiments

In its simple version, the interaction between ligands and small-peptides or short nucleic acids sequences, may also be studied from the receptor's perspective by monitoring their ^{1}H CSPs in standard two-dimensional experiments (TOCSY, NOESY) using an unlabeled receptor. In such cases, the assignment of the receptor's ^{1}H NMR signals can be achieved with the standard strategies used for spectral assignment of small ligands, using a combination of TOCSY and NOESY experiments. Once again, for CSP analysis, two spectra, before and after ligand addition should be recorded.

Obviously, the NOESY spectra also provide information on the conformation and the 3D arrangement of the receptor in both states, free and bound, allowing the observation of conformational changes at the molecular recognition event, if any, and the possible intermolecular contacts.

1.1.4.2.3. Relaxation

As mentioned before, relaxation properties are characteristic of each type of molecule and depend on the global motion rotational correlation time. However, in biomacromolecules, due to the different local dynamics (folding-unfolding, loops, amino acids side chain mobility), local differences may appear that affect their relaxation behavior.

The Lipari and Szabo model, also dubbed model free approach, was developed in the 80s to correlate NMR relaxation data to protein mobility and has been widely accepted. The assumptions of the model free approach are consistent to describe the dynamics of globular isotropic proteins.[21] The employed NMR parameters with dynamic information are the longitudinal relaxation rate (R_1), the transversal relaxation rate (R_2) and the heteronuclear NOE effect (hetNOE). The tendency of these NMR parameters over time is employed for the qualitative interpretation of the protein motions. These parameters describe the mobility of each N-H vector of the amino acids of the backbone. The protein regions with a high mobility are encoded by small R_1 ($R_1=1/T_1$) and R_2 ($R_2=1/T_2$) values and negative hetNOE while motion restricted amino acids display positive hetNOE, large R_2 and small R_1.

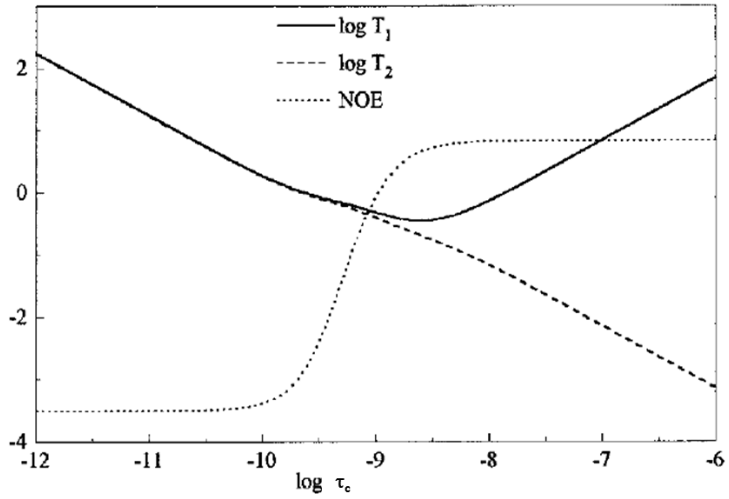


Figure 11. Representation of the variation of T_1 ($1/R_1$), T_2 ($1/R_2$) and hetNOE.

Experimentally, the dynamic study of a protein is expensive. It is a material-consuming process, with a high cost because of the need of ^{15}N labeled protein. It is also time-consuming, since NMR relaxation experiments require long time for the acquisition of long sets of spectra for the case of R_1 and R_2. In addition, for an efficient analysis of the protein dynamics, the NMR relaxation data should be acquired at different magnetic fields, thus increasing the experimental time.

1.2. Key interactions in molecular recognition events

The nature of molecular recognition of biomolecules, such as proteins, carbohydrates or DNA, resides in the existence of weak intermolecular forces between the species. In fact, two major sets of intermolecular forces have been described: polar and non-polar interactions. Nevertheless, other mediated intermolecular interactions may happen.

1.2.1. Polar interactions: Hydrogen bonding[55,56]

The formation of a hydrogen bond (HB), in general, involves the presence of a hydrogen bond donor, characterized by the presence of a hydrogen covalently bonded to an electronegative atom, and a hydrogen acceptor, also usually an electronegative atom. This atom provides a lone pair of electrons to be shared with the donated hydrogen that accommodates the electron density on its valency layer. Frequently, electronegative atoms like O or N atoms may act on both sides of the hydrogen bond, as hydrogen donors when they have covalently bonded hydrogen. Together with fluorine, they may also act as acceptors when they provide a lone pair of electrons to be shared with the donated proton.

Because of the prevalence of these atoms (O, N and H) in biological systems, HB plays an important role in molecular recognition contributing not only to affinity but also to selectivity. HB is the most evident in carbohydrate recognition by proteins due to the presence of numerous –OH groups in saccharides with simultaneous behavior as donor and acceptor that may establish HB with amines and carbonyl groups from the protein backbone and side chains of polar residues (figure 11).

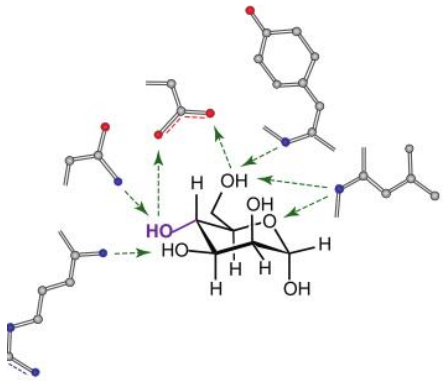


Figure 12. Hydrogen bonding representation in carbohydrate recognition. Adapted from Gabius *et al*.[56]

1.2.2. Non-polar interactions[55–57]

Apart from hydrogen bonding, non-polar interactions also contribute to the molecular recognition between biochemical species. They are of special interest when aromatic protein residues are involved in the recognition event. For instance, the non-polar CH groups of a ligand tend to pack with adjacent aromatic rings establishing CH-pi interactions. The geometry condition for CH-pi interactions requires the presence of CH groups perpendicular to the aromatic ring where the CH-pi system stabilize the complex formation by entropic and enthalpic contributions. Since the stabilizing energy of one CH-pi interaction is rather low, (CH-pi interaction ca.= 1 kcal/mol), the simultaneous CH-pi interaction of various CH groups could finally drive the molecular recognition. Indeed, CH-pi interactions have been deeply described in carbohydrate recognition[58,59] (figure 12).

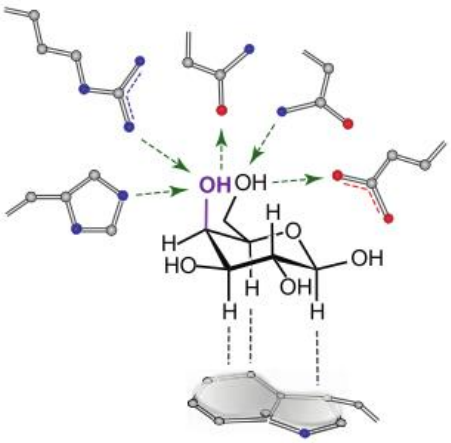


Figure 13. CH-pi interaction representation in carbohydrate recognition. Adapted from Gabius *et al*.[56]

1.2.3. Other interactions[55,56]

Due to the complexity of the biochemical systems in molecular recognition, other polar and non-polar intermolecular interactions may take place. Some of them involve third partners, such as ions or water molecules. Water mediated interactions take place when a water molecule establishes HB simultaneously with both partners, ligand and receptor. The nature of this interaction resides in the presence of solvating water molecules in the binding pocket that enhances the ligand binding mediating as a "bridge" between the ligand and receptor.

Similarly, ion mediated intermolecular interactions take the advantage of electrostatic forces established between both partners and one ion present at or close to the binding site. As water mediated interactions, the ion may act as a bridge between the partners and/or to stabilize the amino acid side chains in the proper orientations to bind the ligand (figure 13).

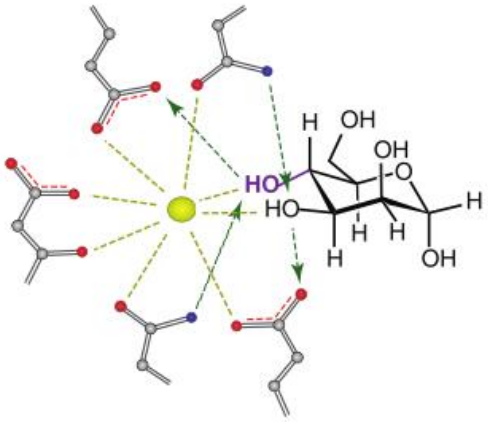


Figure 14. Ion mediated interaction in carbohydrate recognition. Adapted from Gabius *et al*.[56]

In summary, molecular recognition involves the cooperative contributions of several intermolecular forces to form stable supramolecular receptor-ligand complexes.

1.3. Ligands and receptors

All types of molecules of biological interest, proteins, nucleic acids, carbohydrates, vitamins, drugs, etc., participate in one way or another as recognition partners. Therefore, the study of these processes at atomic molecular level is of paramount importance for understanding their mechanism and functionality and the pathological consequences derived of their malfunction. In a further step, this knowledge would drive the basis for drug design strategies to control and modulate such processes in a therapeutic context. In the present thesis, some examples of

ligand-receptor partners, carbohydrates and their protein receptors as lectins and antibodies, natural products that bind tubulin, an essential protein of cell cytoskeleton, synthetic compounds and peptides that can bind to nucleic acids or control its transcription, have been studied by the NMR methodologies introduced above.

1.3.1. Carbohydrates

During long time, carbohydrates (glycans, sugars, saccharides) were exclusively considered as the first source of energy for living organisms while the study of further capabilities was not relevant. Large efforts have been put on the last decades to understand transcriptional and translational processes. However, little attention was initially paid to post-transcriptional processes. Nevertheless, after the observation of carbohydrate in membranes, other aspects started to be scrutinized. Therefore, the discovery of the glycocalyx has motivated the study of their biological functions, thus boosting advances in glycosciences including glycomics.

It is evident today that carbohydrates, when conjugated as glycoproteins or glycolipids at the glycocalix, play an important role in numerous life processes and also transmit important biological information. Carbohydrates at the glycocalix are involved in molecular recognition processes by establishing interactions with antibodies, proteins, viruses, cells, bacteria… (figure 14) that trigger a variety of intracellular and extracellular processes of special physiological and pathological interest.

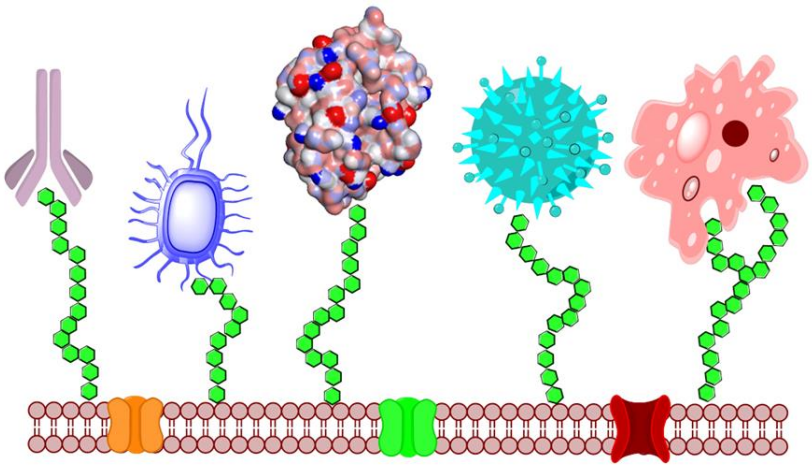


Figure 15. Cell membrane carbohydrates and their interaction with antibodies, bacteria, proteins, virus and cells.

1.3.1.1. Monosaccharides

Carbohydrates constitute a large group of biomolecules built by monosaccharides. In nature, different monosaccharides exist with different number of carbon atoms. All of them share similar structural features. Monosaccharides are composed by poly-hydroxylated carbon chains, usually from five to seven carbon atoms, with a carbonyl group ketone (ketoses) or aldehyde (aldoses). The high diversity and complexity of carbohydrates may be even noticed at the monosaccharide level due to the vast number of possible monosaccharides. However, this structural variability is even increased when the monosaccharide is studied in solution because of their cyclization event.

For the case of hexoses, for instance glucose (Glc), two different heterocycles could be formed in solution: 5-membered ring (furanose) or 6-membered ring (pyranose) (figure 15). Moreover, the cyclization introduces a new source of variability of special interest in pyranoses related with the flexibility of the 6-membered ring and the spatial orientation of the functional groups in either axial or equatorial orientation.[60,61] Therefore different conformation may be adopted in fast equilibrium among them. In the case of Glc,[62] the most populated conformation is the chair form $^4C_1$ since the all the hydroxyl groups are in equatorial minimizing the steric hindrance with axial hydrogens (figure 16). The molecular recognition features of carbohydrates have been largely explored and sugar conformation was found critical for the complex formation.[63]

Even though, in a 6-membered ring, the minimum of energy is acquired by the accommodation of bulky groups in equatorial orientation, pyranoses experienced a stabilization of the axial orientation of the hydroxyl of the hemiacetalic moiety at the anomeric position (carbon 1 in the Glc example in figure 17) dubbed anomeric effect.

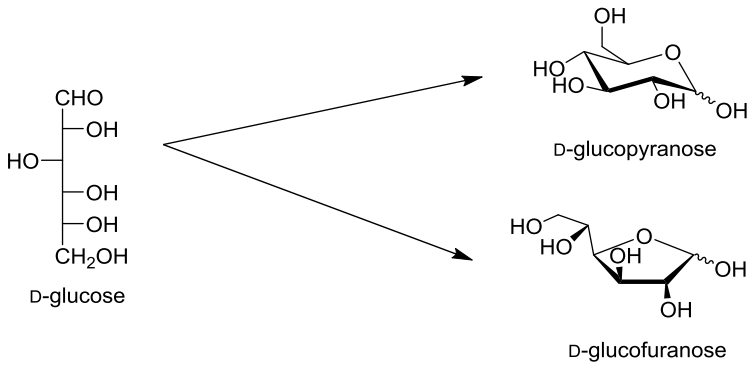


Figure 16. Schematic representation of glucose (Glc). On the left: open chain form. On the right: possible cyclizations in solution.

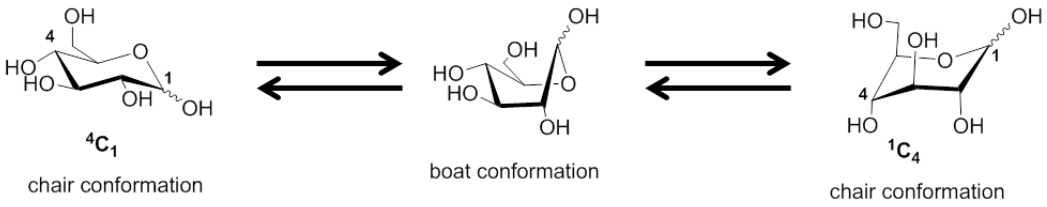


Figure 17. Conformational equilibrium of glucopyranose (Glc*p*).

1.3.1.2. The anomeric effect[64–68]

The anomeric effect was firstly postulated in the 1950s by Edward[69] and Lemieux and it has remained a controversial topic in science. The anomeric effect describes the stabilization of the axial electronegative substituent at the anomeric position due to orbital and electronic factors produced by the presence of a cyclic oxygen atom. The tendency of the anomeric -OH or electronegative substituent to adopt axial orientation (α-anomer) rather than equatorial (β-anomer) is in contrast to what would be expected based on steric factors (figure 18). However, modelling and computational approaches have permitted the understanding of this phenomenon.[70,71]

The stabilization of the α-anomer is produced by the hyperconjugative effect between the lone pair orbitals from the endocyclic oxygen and the s* orbital of the C-X bond with the exocyclic substituent. In the axial configuration, this antibonding orbital displays the proper orientation to accommodate lone pair of electrons from the endocyclic oxygen.[66] This orbital conjugation is not possible in the equatorial configuration of the β-anomer.

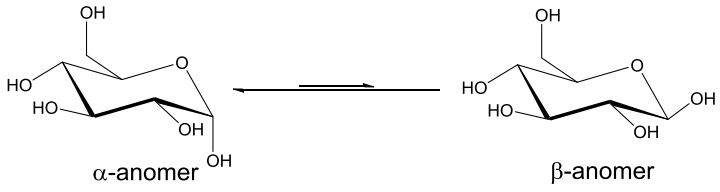


Figure 18. Equilibrium between Glc*p* anomers. The intermediate open form is not represented.

1.3.1.3. Oligo and polysaccharides

The association between monosaccharides (glycosylation) results in more complex biomolecules, oligo- and polysaccharides. The linkage between two monosaccharide units receives the name of glycosidic linkage and present several features that make it unique. The

different conformations adopted around the glycosidic linkage are defined by the torsion angles Φ and Ψ . Additionally in the case of 1-6 linkages there is one additional torsion, ω (figure 19). The geometry of the glycosidic linkage has also steric, electronic and orbital factors (the exo-anomeric effect)[68,72,73] that favors the adoption of certain range of angles.[74]

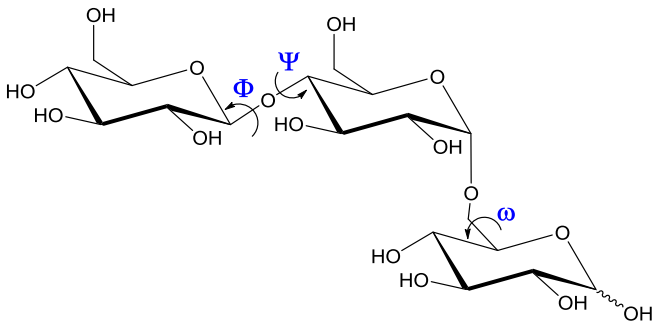


Figure 19. Representation of the glycosidic linkage angles.

If the variability of monosaccharides is enormous, the variability of oligosaccharides is much higher and increases exponentially with the number of units. From the linear disaccharide, an additional third unit could result in a branched or linear trisaccharide with its concomitant level of complexity. Therefore, over other biomolecules, the number of possible combinations of oligosaccharides gives to carbohydrates the ability to encode huge biological information with few monosaccharide units.[75] Moreover, carbohydrate modifications have been also observed in nature with critical consequences. For instance, blood group A and B (figure 20) compatibility is driven by a single modification on the oligosaccharide, the presence or absence of a N-acetyl group in the terminal galactose (Gal) results in the rejection of the blood type.[76,77]

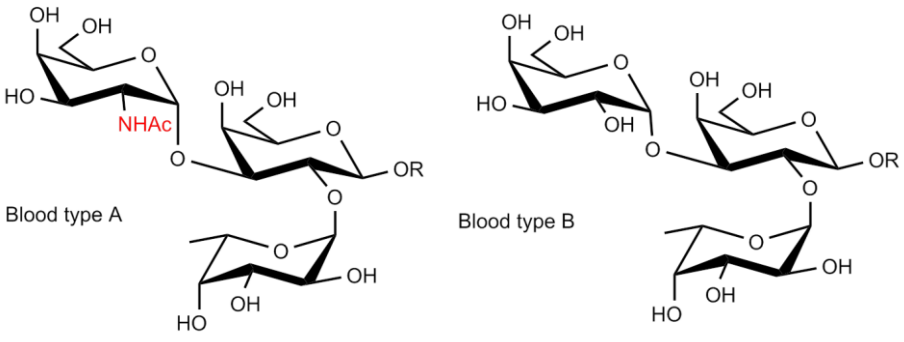


Figure 20. Chemical structures of the oligosaccharides of blood group types A and B.

In addition, carbohydrates, from monosaccharide units to polysaccharides, may be also conjugated with other biomolecules such lipids or proteins. Actually, more than half of human proteins display glycosylation in post-translational modifications constituting glycoproteins. Glycosylated lipids form glycolipids.[78,79]

1.3.2. Proteins

Proteins are the most common receptors in the biological molecular recognition processes. Carbohydrates and ligands are recognized by biochemical partners belonging to different protein families such as enzymes, lectins, antibodies or others.

Enzymes are a class of proteins involved in biochemical transformations. In fact, the main metabolic transformations of sugars are mediated by enzymes that transfer monosaccharides to different scaffolds (glycotransferases) or destroy glycosidic linkages to obtain carbohydrate subunits (glycosidases).[80] Additionally, a great variety of other types of enzymes (oxidoreductases, isomerases, kinases, acylases, etc) are responsible of further chemical modifications of the monosaccharide units. One of the key principles of the enzymes is their precise substrate selectivity and conversion efficiency that has been exploited in-vitro in carbohydrate synthesis facilitating the chemo-enzymatic synthesis of complex oligosaccharide structures.[81–85]

Lectins are a large family of proteins from different sources that specifically recognize carbohydrates without producing any transformation on them. Historically, the term “lectin” derives from the latin verb “legere” that means “to select” and was proposed by W. C. Boyd in 1954[86] for naming some specific blood agglutinins from vegetal seeds and specific of blood groups. Nowadays, “lectin” is applied, in general, to proteins (excluding antibodies and enzymes) that interacts reversibly with saccharides.[87]

Finally, antibodies consist in proteins from the immune system important in the defense of the organism against pathogens and other external aggressions and distinguish “self” from “non-self” structures. Many pathogens and aberrant cells present characteristic oligo- or polysaccharide structures in their cellular envelopes.[88,89] The immune system is able to interpret carbohydrate structures as cell signatures and, hence, to synthesize specific antibodies against non-self-structures to protect the organism.[90,91] Interestingly, as mention above, a recognition process of special relevance for human species is the identifications of blood groups ABO (figure 20) that is mediated by carbohydrate-antibody interactions.[76,77,92]

Proteins are the main executors of life processes. An alteration of the protein function often triggers pathological disorders, misfunction or over- or under-expression. However, different therapeutic approaches based on ligand binding to these therapeutics targets have been developed to inhibit or promote protein activity or to inhibit or promote protein expression. Thus, proteins are, in essence, the main concern in the molecular recognition applied to medicinal chemistry. The further knowledge of the protein molecular recognition features and structure led to the better design of inhibitor and binders. In this context, other proteins that are involved in important cell and life processes such the case of tubulin, which plays a main role in mitosis, has been validated as antitumor therapeutic target.[93]

1.3.2.1. Lectins

As mentioned above, lectins belong to a carbohydrate binding protein family distinct from antibodies and enzymes. Lectins are glycan specific receptors that present a structural diversity and are ubiquitous in nature. Indeed, lectins are present in virus and microbes, plants, animals and humans (figure 21). Due to the diversity of the lectins, many different classifications attending to carbohydrate specificity, structure or functions have been proposed.

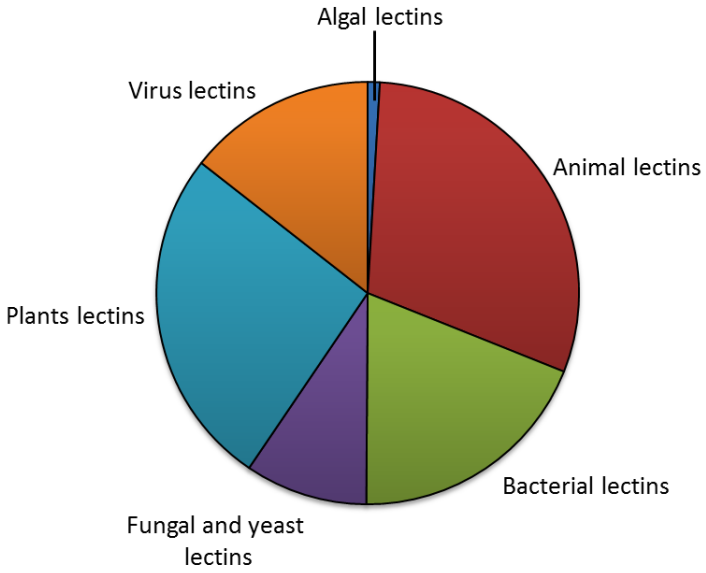


Figure 21. Lectin distribution over different species.

Among the lectin family, certain amino acids have been found at the carbohydrate binding site that drives carbohydrate specificity. This applies to aromatic residues that establish CH-pi interactions with the hydrophobic face of the sugar and arginine, aspartate and glutamate

residues which may establish bidentate HB interactions with hydroxyl groups of the saccharide.[94]

Galectins, a family of 15 lectins with affinity to β-galactose[95,96] containing oligosaccharides and a basic structural β-sandwich fold, display a conserved carbohydrate recognition domain (CRD) of around 130 amino acids where a central tryptophan (Trp) residue is the key feature for carbohydrate interaction.[97] These structural features differentiate galectins from other β-sandwich lectins such as calnexin or pentraxins. In galectins, the CRD may be found in three types of structural contexts, proto- (non-covalent association of CRDs), chimera- (CRD connected to an N-terminal tail) and tandem repeat type (CRDs linked by a peptidic sequence) arrangements.[98–100] Remarkably, all galectins are confined in proto- and tandem repeat type except for galectin 3 (gal-3) which is the unique chimera type galectin. Galectins often exists as oligomers of CRDs so they are multivalent proteins able to bind to several carbohydrate epitopes, for instance, gal-3 forms pentamers in the presence of multivalent glycan ligands.[101,102] In addition, galectin functions are related to their tendency to dimerize or oligomerize comprising a large number of cellular activities such as cell-cell interaction, cell-pathogen interaction or signal transduction through receptor clustering (figure 22).[102–105]

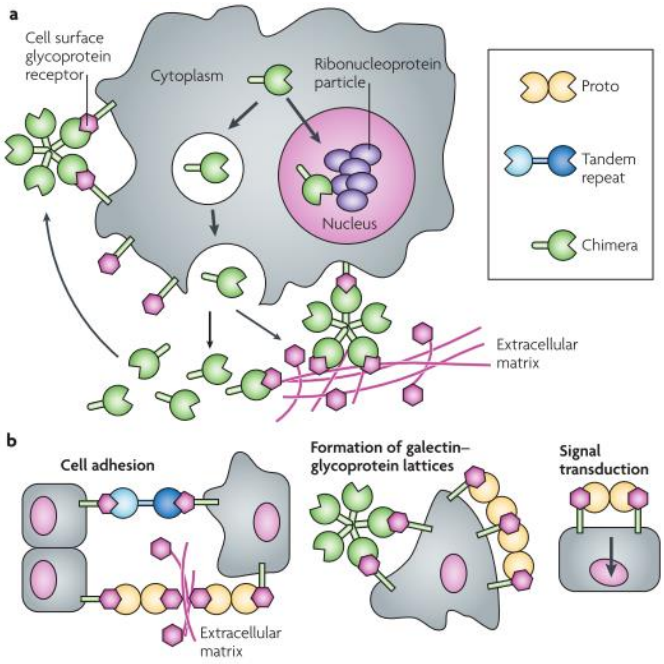


Figure 22. Intra- and extra- cellular functions of galectins. Adapted from Vasta [106]

Moreover, many galectins exhibit dual localization[107] in the cell being found in both intracellular and extracellular matrix displaying different activities. Gal-3 is present in the nucleus and cytoplasm of the cell as well as in the extracellular matrix. Depending on the

cellular localization gal-3 displays different activities most of them related with cell proliferation in the immune and inflammatory responses. Of note, gal-3 induces apoptosis in T cells[108] but anti apoptotic activity has been reported in macrophages[109] and neutrophils.[110] Thus, endogenous gal-3 in T cells is anti-apoptotic, but extracellular gal-3 secreted by other cells may kill T cells through binding to T cell surface glycoproteins. Overall, current information suggest that inhibitors of gal-3 may be useful for treatment of inflammatory diseases.[111]

Viral lectins are involved at the early infection events by recognizing oligosaccharides at the surface of their targeted cells. In fact, that is the case of the hemagglutinin (HA) of the influenza virus directly involved in influenza infection. The influenza HA consists in a family of 16 structural related lectins specific for terminal sialic acid glyconjugates. The HA is built by the association of three identical monomers with a carbohydrate binding site in each monomer; therefore, the HA unit (trimer) presents mulitvalency towards sialic acid. The carbohydrate binding site displays a number of conserved amino acids as well as three conserved elements of secondary structure, the 130- and 220- loops and the 190-α-helix (figure 23).

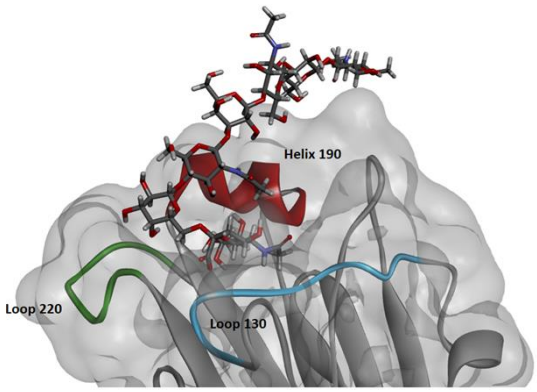


Figure 23. Carbohydrate binding site of the influenza hemagglutinin H1.

Since influenza viruses may affect different species such as avians and mammals, including humans, structural and specificity insights have been obtained for a better understanding of the infection. Indeed, the glycosidic linkage of the terminal sialic acid to the glycoconjugate determines the infection of particular specie. While avian influenza HA strains display binding towards α(2-3) sialic acid linked to galactose, the human HA strain is specific of α(2-6). These subtle differences are the key of the transfection between species. The affinity of the HA, as occurs in most of lectins, is rather low. Consequently the tight binding of the HA towards sialic acid is mediated by its multivalency with the simultaneous binding of sialosides.[112,113] The

influence of the cooperativity on effective binding affinity has encouraged studies of the inhibition of receptor binding by modified sialosides presented as polyvalent inhibitors.[114–116]

1.3.2.2. Antibodies

Antibodies are a large class of proteins directly involved in the immune response against antigens. The antibody's activity comprises the binding to their specific target antigens and the activation of the other immune cells and molecules triggering the immune response. The exquisite specificity and high affinity of the antibodies is the key for immunity and has been exploited for immunohistochemistry, diagnostics or biomedical research arising as a valuable tool in science.

Antibodies are composed by small globular protein domains belonging to immunoglobulin family. The antigen binding fragment (Fab) is composed by two identical polypeptide chains of ~500 amino acids (heavy chain, H) attached by disulfide bridges to two identical polypeptide light chains of ~250 amino acids (light chain, L). At the same time, H and L chains are composed by a variable N-terminal region (V) respectively and one constant C-terminal region (C) for the L chain and three for the H chain (figure 24). The high variability of the variable domains, where antigen binding occurs, is the key feature to the immense number of different antibody sequences already described and their specificities.

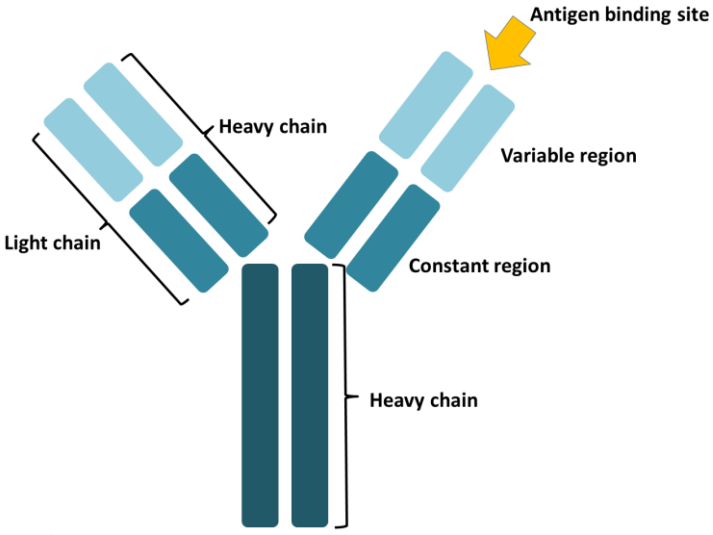


Figure 24. Schematic representation of an immunoglobulin.

The understanding of the antibody structural motifs has permitted the development of antibody engineering for the expression of tailored-made monoclonal antibodies useful for vaccines, immunoassays and therapeutics.

Over a variety of antigens, antibodies recognize carbohydrate-based antigens from glycoproteins or glycolipids at the cell glycocalix with high specificity. For instance, the recognition by antibodies of the α-Gal epitope of the xenoantigen, present in xenografts triggers, within minutes, the hyperacute rejection of the organ.[117–119] The knowledge of binding epitopes of these glycan antigens is an essential input for the very active field in saccharide-based vaccines development.[120]

1.3.2.3. Other protein receptors: Tubulin

Tubulin is the structural building block of microtubules (MT), the main component of cellular cytoskeleton. It is essential for cell processes such as cell division, vesicle transport or cell shape[93,121,122] and consequently an attractive therapeutic target for anticancer drugs.[93] In fact, numerous anticancer drugs have been developed targeting MT and interfering in microtubule dynamics inducing the cell arrest and apoptosis.[123] In addition, MT targeted anticancer drugs have been approved for cancer treatment and are in current clinical use (Taxol, Vinblastine) while others are or have been in advanced clinical trials. The mechanism of tubulin binding agents (TBAs) is focused on the stabilization or destabilization of the MT.

MTs are long tube shaped protein polymers composed by the polymerization of tubulin heterodimers formed, at the same time, by two subunits: α-tubulin and β-tubulin (figure 25). The MT dynamic behavior, polymerization and de-polymerization, is defined as dynamic instability and mediated by GTP/GDP exchange through tubulin subunits (α and β).[124] Indeed, the α-tubulin subunit is always bound to GTP, acquiring a conformation suitable for MT polymerization, whereas the β-tubulin subunit can be bound either to GTP or GDP, favorable for MT polymerization or de- polymerization, respectively (figure 25).[125,126] Consequently, two terminal motifs may be distinguished in a MT, the dubbed (+) end (terminated by β-tubulin) that grows and shortens more rapidly than the (-) end, which is terminated by α-subunit. In humans, a total of nine α- and nine β- tubulin isotypes have been identified, that differs on the C-terminal sequence.[127] Different post-translational modifications may affect the dynamic properties of MT and its modulation.[128]

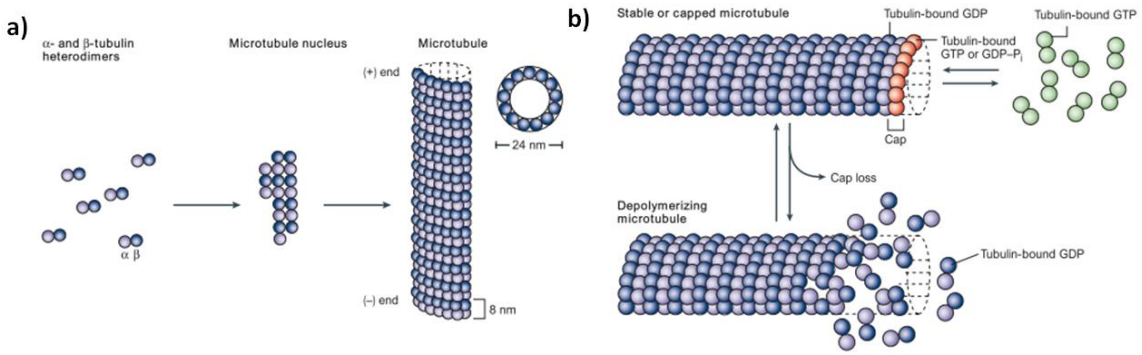


Figure 25. Panel A:Polymerization of MTs. Panel B: MTs dynamics. Adapted from Jordan *et al*. [93]

MT dynamics is carefully regulated in cells and permit that MT performs multiple functions during the cell life-time.[93,122,129] During the interphase, MTs mediate in intracellular trafficking and are responsible of cell shape. During mitosis, MT dynamics dramatically increase for the attachment, separation and segregation of chromosomes.[122] The alteration of MT dynamics is the key principle of action of MT targeted drugs[123] such as taxol-like molecules, which block mitosis and kill tumor cells.[93]

Of note, the microtubule contains two microtubule stabilizing agent binding sites in β-tubulin subunit,[130] the luminal and the pore site. While the luminal site is present in the tubuin heterodimer and in MT, the pore site appears at the interface of heterodimers and therefore is only present in MT.[130] The case of the reference taxane anti-tumor drug paclitaxel is thought to gain access to its binding site by diffusion from one site to the other.[125] Epothilones, a family of polyketide macrolactone natural compounds identified in soil bacteria, are currently under study as alternative to taxanes as antitumor drugs targeting the luminal site.[131]

1.3.3. Nucleic acids: DNA

DNA encodes the genetic information needed for life processes. Thus, an aberration on the DNA sequence or on its translation may trigger lethal consequences. Hence, DNA has been considered the traditional target for medicinal chemistry to treat cancer for many years.

The 3D structure elucidation of double-stranded DNA in 1953 by Watson and Crick[132] was a revolutionary challenge in biology awarded with the Nobel Award. However, apart from the double stranded DNA, other supramolecular DNA arrangements such as triple helix DNA[133,134] and quadruplexes[135] also naturally occurs. Many natural and synthetic products have been developed that associate with the double-stranded DNA helix with potential therapeutic applications.[134,136–138]

From the molecular recognition point of view, DNA transcription may be considered as a great example of selective recognition by proteins where numerous protein systems (transcription factors, TFs) selectively recognize consensus DNA sequences to start the biological processes of the central dogma of the biology.[139]

Transcription factors (TFs) are specialized proteins that participate in the regulation of gene expression by binding to key DNA regulatory sequences, and thereby promoting or inhibiting the assembly of the transcriptional complex.[140] TFs display a modular structure built by peptide domains with a variety of functions from DNA binding to stimulatory effect on transcription.[140] TFs are often classified on the basis of DNA binding structural elements. Of note, the basic DNA binding domains, for instance C/EBP family,[141,142] are often found in association with leucine zipper motifs and only bind to DNA once TFs dimers are formed.[143]

From the medicinal point of view, TFs may be considered as interesting therapeutic targets since many TFs have been implicated in human diseases (including cancer), inflammatory or heart diseases. Thus, TFs mimetics have been designed for gene medicine[144–146] and hence control the DNA transcription. The key feature of TFs mimetics is the achievement of sequence selective mimetics with significant affinity; therefore, the development of sequence selective TFs mimetics is still challenging with an enormous potential in gene medicine.

1.4. References

(1) Pierce, M. M.; Raman, C. S.; Nall, B. T. *Methods* **1999**, *19* (2), 213.
(2) Jelesarov, I.; Bosshard, H. R. *J. Mol. Recognit.* **1999**, *12* (1), 3.
(3) Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J. *Nature* **2007**, *448* (7151), 325.
(4) Liu, F.; Liu, X.; Ng, S.-C.; Chan, H. S.-O. *Sensors Actuators B Chem.* **2006**, *113* (1), 234.
(5) Liu, Y.; Yu, X.; Zhao, R.; Shangguan, D.-H.; Bo, Z.; Liu, G. *Biosens. Bioelectron.* **2003**, *19* (1), 9.
(6) Karlsson, R. *J. Mol. Recognit.* **2004**, *17* (3), 151.
(7) Terada, Y.; Seto, H.; Hoshino, Y.; Murakami, T.; Shinohara, S.; Tamada, K.; Miura, Y. *Polym J*. The Society of Polymer Science, Japan February 2017, pp 255–262.
(8) Stsiapanava, A.; Samuelsson, B.; Haeggström, J. Z. *Proc. Natl. Acad. Sci.* **2017**.
(9) Fellner, M.; Desguin, B.; Hausinger, R. P.; Hu, J. *Proc. Natl. Acad. Sci.* **2017**, *114* (34), 9074.
(10) Unione, L.; Galante, S.; Díaz, D.; Cañada, F. J.; Jiménez-Barbero, J. *Med. Chem. Comm* **2014**, *0*, 1.
(11) Mayer, M.; Meyer, B. *Angew. Chem. Int. Ed.* **1999**, *38* (12), 1784.
(12) Marcelo, F.; Dias, C.; Martins, A.; Madeira, P. J.; Jorge, T.; Florêncio, M. H.; Cañada, F. J.; Cabrita, E. J.; Jiménez-Barbero, J.; Rauter, A. P. *Chem. Eur. J.* **2013**, *19* (21), 6641.
(13) Gimeno, A.; Reichardt, N.-C.; Cañada, F. J.; Perkams, L.; Unverzagt, C.; Jiménez-Barbero, J.; Ardá, A. *ACS Chem. Biol.* **2017**, *12* (4), 1104.

(14) Dyson, H. J.; Wright, P. E. *Chem. Rev.* **2004**, *104* (8), 3607.
(15) Canales, A.; Nieto, L.; Rodríguez-Salarichs, J.; Sánchez-Murcia, P. A.; Coderch, C.; Cortés-Cabrera, A.; Paterson, I.; Carlomagno, T.; Gago, F.; Andreu, J. M.; Altmann, K. H.; Jiménez-Barbero, J.; Díaz, J. F. *ACS Chem. Biol.* **2014**, *9* (4), 1033.
(16) Ardá, A.; Bosco, R.; Sastre, J.; Cañada, F. J.; André, S.; Gabius, H.-J.; Richichi, B.; Jiménez-Barbero, J.; Nativi, C. *Eur. J. Org. Chem.* **2015**, *2015* (31), 6823.
(17) Pomin, V. H. *Glycobiology*. 2014, pp 991–1003.
(18) Bräuniger, T.; Jansen, M. *Zeitschrift für Anorg. und Allg. Chemie* **2013**, *639* (6), 857.
(19) Truflandier, L. A. Poeppelmeier, K. B. T.-C. I. C. I. I. (Second E., Ed.; Elsevier: Amsterdam, **2013**; pp 381–406.
(20) Gunther, H. In *NMR Spectroscopy: Basic Principles, Concepts and Applications in Chemistry*; **2013**; pp 13–28.
(21) Gunther, H. *NMR Spectroscopy: Basic Principles, Concepts and Applications in Chemistry*; Wiley-VCH, **2013**.
(22) Claridge, T. D. . *Protein – Ligand Screening by NMR*; **2016**.
(23) Claridge, T. D. W. *Diffusion NMR Spectroscopy*; **2016**; Vol. 27.
(24) Price, W. S. *Concept. Magn. Reson.* **1998**, *10* (4), 197.
(25) Brand, T.; Cabrita, E. J.; Berger, S. *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *46* (4), 159.
(26) Yan, J.; Kline, A. D.; Mo, H.; Zartler, E. R.; Shapiro, M. J. *J. Am. Chem. Soc.* **2002**, *124* (34), 9984.
(27) Fielding, L. *Prog. Nucl. Magn. Reson. Spectrosc.* **2007**, *51* (4), 219.
(28) Fernandez-Alonso, M. del C.; Alvaro Berbis, M.; Canales, A.; Arda, A.; Javier Canada, F.; Jimenez-Barbero, J. In *New Applications of NMR in Drug Discovery and Development*; The Royal Society of Chemistry, **2013**; pp 7–42.
(29) Zerbe, O. *BioNMR in Drug Research*; **2003**.
(30) In *Structure-Based Drug Discovery: An Overview*; Hubbard, R. E., Ed.; The Royal Society of Chemistry, **2006**; pp 97–141.
(31) Erlanson, D. A. Davies, T. G., Hyvönen, M., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, **2012**; pp 1–32.
(32) Davies, Thomas G.; Hyvönen, M. *Fragment-Based Drug Discovery and X-Ray Crystallography*; **2012**; Vol. 317.
(33) Andricopulo, L. G. F. and A. D. *Curr. Top. Med. Chem.* **2017**, pp 2260–2270.
(34) Harner, M. J.; Frank, A. O.; Fesik, S. W. *J. Biomol. NMR* **2013**, *56* (2), 65.
(35) Mayer, M.; Meyer, B. *Angew. Chemie Int. Ed.* **1999**, *38* (12), 1784.
(36) Meyer, B.; Peters, T. *Angew. Chem. Int. Ed.* **2003**, *42* (8), 864.
(37) Mayer, M.; Meyer, B. *J. Am. Chem. Soc.* **2001**, *123* (25), 6108.
(38) Enríquez-Navas, P. M.; Marradi, M.; Padro, D.; Angulo, J.; Penadés, S. *Chem. – A Eur. J.* **2011**, *17* (5), 1547.
(39) Ribeiro, P. J.; Diercks, T.; Jiménez-Barbero, J.; André, S.; Gabius, H.-J.; Cañada, J. F. *Biomolecules* . **2015**, 5 (4), 3177
(40) Dalvit, C.; Fogliatto, G.; Stewart, A.; Veronesi, M.; Stockman, B. *J. Biomol. NMR* **2001**, *21* (4), 349.
(41) Dalvit, C.; Pevarello, P.; Tato, M.; Veronesi, M.; Vulpetti, A.; Sundstrom, M. *J. Biomol. NMR* **2000**, *18* (1), 65.
(42) Otting, G.; Liepinsh, E. *Acc. Chem. Res.* **1995**, *28* (4), 171.
(43) Balaram, P.; Bothner-By, A. A.; Breslow, E. *J. Am. Chem. Soc.* **1972**, *94* (11), 4017.
(44) Balaram, P.; Bothner-By, A. A.; Dadok, J. *J. Am. Chem. Soc.* **1972**, *94* (11), 4015.
(45) Manzenrieder, F.; Frank, A. O.; Kessler, H. *Angew. Chem. Int. Ed.* **2008**, *47* (14), 2608.
(46) Dolbier, W. R. *J. Fluor. Chem.* **2005**, *126* (2), 157.
(47) Kirk, K. L. *J. Fluor. Chem.* **2006**, *127* (8), 1013.
(48) Purser, S.; Moore, P. R.; Swallow, S.; Gouverneur, V. *Chem. Soc. Rev.* **2008**, *37* (2), 320.
(49) Hagmann, W. K. *J. Med. Chem.* **2008**, *51* (15), 4359.

(50) Dalvit, C.; Flocco, M.; Veronesi, M.; Stockman, B. J. *Combinatorial Chemistry & High Throughput Screening*.**2002,** pp 605–611.
(51) Dalvit, C.; Ardini, E.; Flocco, M.; Fogliatto, G. P.; Mongelli, N.; Veronesit, M. *J. Am. Chem. Soc.* **2003**, *125* (47), 14620.
(52) Dalvit, C.; Piotto, M. *Magn. Reson. Chem.* **2017**, *55* (2), 106.
(53) Diercks, T.; Ribeiro, J. P.; Cañada, F. J.; André, S.; Jiménez-Barbero, J.; Gabius, H.-J. *Chem. Eur. J.* **2009**, *15* (23), 5666.
(54) Berbís, M. Á.; Canales, Á.; Sastre-Martinez, J. *Advanced NMR Techniques: Defining Carbohydrate Structures and Ligand–Receptor Interactions*, **2014**, 121.
(55) del Carmen Fernández-Alonso, M.; Díaz, D.; Berbis, M. Á.; Marcelo, F.; Cañada, J.; Jiménez-Barbero, J. *Curr. Protein Pept. Sc.* **2012**, *13* (8), 816.
(56) Gabius, H. J.; André, S.; Jiménez-Barbero, J.; Romero, A.; Solís, D. *Trends Biochem. Sci.* **2011**, *36* (6), 298.
(57) Fernández-Alonso, M. C.; Cañada, F. J.; Jiménez-Barbero, J.; Cuevas, G. *J. Am. Chem. Soc.* **2005**, *127* (20), 7379.
(58) Asensio, J. L.; Ardá, A.; Cañada, F. J.; Jiménez-Barbero, J. *Acc. Chem. Res.* **2013**, *46* (4), 946.
(59) Muraki, M. *Protein Peptide Lett*. **2002**, 9 (3), 195.
(60) Sattelle, B. M.; Bose-Basu, B.; Tessier, M.; Woods, R. J.; Serianni, A. S.; Almond, A. *J. Phys. Chem. B* **2012**, *116* (22), 6380.
(61) Ronnols, J.; Manner, S.; Ellervik, U.; Widmalm, G. *Org. Biomol. Chem.* **2014**, *12* (40), 8031.
(62) Biarnés, X.; Ardèvol, A.; Planas, A.; Rovira, C.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2007**, *129* (35), 10686.
(63) Glaudemans, C. P. J.; Lerner, L.; Daves, G. D.; Kovac, P.; Venable, R.; Bax, A. *Biochemistry* **1990**, *29* (49), 10906.
(64) Rauter, A. P.; Lindhorst, T. *Carbohydrate Chemistry*; Rauter, A. P., Lindhorst, T., Eds.; Royal Society of Chemistry, **2003**.
(65) Gerber-Lemaire, S.; Vogel, P. In *Carbohydrate Chemistry: Volume 35*; The Royal Society of Chemistry, **2009**; 35, 13.
(66) Freitas, M. P. *Org. Biomol. Chem.* **2013**, *11* (17), 2885.
(67) Juaristi, E.; Cuevas, G. *Tetrahedron* **1992**, *48* (24), 5019.
(68) Tvaroŝka, I.; Bleha, T. *Adv. Carbohydr. Chem. Biochem.* **1989**, *47*, 45.
(69) J. T., E. *Chem. Ind.* **1955**, 1102.
(70) Huang, Y.; Zhong, A.-G.; Yang, Q.; Liu, S. *J. Chem. Phys.* **2011**, *134* (8), 84103.
(71) Lii, J.-H.; Chen, K.-H.; Durkin, K. A.; Allinger, N. L. *J. Comput. Chem.* **2003**, *24* (12), 1473.
(72) Thøgersen, H.; Lemieux, R.; Bock, K.; Meyer, B. *Can. J. Chem.* **1981**, *60* (1), 44.
(73) Lemieux, R. U.; Koto, S. *Tetrahedron* **1974**, *30* (13), 1933.
(74) Widmalm, G. *Carbohydr. Res.* **2013**, *378*, 123.
(75) Gabius, H.-J. *Naturwissenschaften* **2000**, *87* (3), 108.
(76) Meloncelli, P. J.; West, L. J.; Lowary, T. L. *Carbohydr. Res.* **2011**, *346* (12), 1406.
(77) Hakomori, S. *Biochim. Biophys. Acta - Gen. Subj.* **1999**, *1473* (1), 247.
(78) Apweiler, R.; Hermjakob, H.; Sharon, N. *Biochim. Biophys. Acta - Gen. Subj.* **1999**, *1473* (1), 4.
(79) Wormald, M. R.; Petrescu, A. J.; Pao, Y.-L.; Glithero, A.; Elliott, T.; Dwek, R. A. *Chem. Rev.* **2002**, *102* (2), 371.
(80) Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B. *Nucleic Acids Res.* **2014**, *42* (D1), 490.
(81) Boltje, T. J.; Buskas, T.; Boons, G.-J. *Nat Chem* **2009**, *1* (8), 611.
(82) Wang, Z.; Chinoy, Z. S.; Ambre, S. G.; Peng, W.; McBride, R.; de Vries, R. P.; Glushka, J.; Paulson, J. C.; Boons, G.-J. *Science* **2013**, *341* (6144), 379.
(83) Totani, K.; Kubota, T.; Kuroda, T.; Murata, T.; Hidari, K. I.; Suzuki, T.; Suzuki, Y.;

Kobayashi, K.; Ashida, H.; Yamamoto, K.; Usui, T. *Glycobiology* **2003**, *13*.
(84) Sauerzapfe, B.; Křenek, K.; Schmiedel, J.; Wakarchuk, W. W.; Pelantová, H.; Křen, V.; Elling, L. *Glycoconj. J.* **2009**, *26* (2), 141.
(85) Sugiarto, G.; Lau, K.; Qu, J.; Li, Y.; Lim, S.; Mu, S.; Ames, J. B.; Fisher, A. J.; Chen, X. *ACS Chem. Biol.* **2012**, *7* (7), 1232.
(86) Boyd, W. C.; Shapleigh, E. *Science.* **1954**, *119* (3091), 419.
(87) Gabius, H. In *The Sugar Code: Fundamentals of Glycosciences*; Wiley-VCH, 2009.
(88) Arthur, C. M.; Cummings, R. D.; Stowell, S. R. *Curr. Opin. Chem. Biol.* **2014**, *18* (Supplement C), 55.
(89) Astronomo, R. D.; Burton, D. R. *Nat. Rev. Drug Discov.* **2010**, *9* (4), 308.
(90) Paulson, J. C.; Blixt, O.; Collins, B. E. *Nat. Chem. Biol.* **2006**, *2* (5), 238.
(91) Schmidt, M. A.; Riley, L. W.; Benz, I. *Trends Microbiol.* **2017**, *11* (12), 554.
(92) Agostino, M.; Sandrin, M. S.; Thompson, P. E.; Yuriev, E.; Ramsland, P. A. *Mol. Immunol.* **2009**, *47* (2), 233.
(93) Jordan, M. A.; Wilson, L. *Nat. Rev. Cancer* **2004**, *4* (4), 253.
(94) Taroni, C.; Jones, S.; Thornton, J. M. **2000**, *13* (2), 89.
(95) Barondes, S. H.; Castronovo, V.; Cooper, D. N. W.; Cummings, R. D.; Drickamer, K.; Felzi, T.; Gitt, M. A.; Hirabayashi, J.; Hughes, C.; Kasai, K.; Leffler, H.; Liu, F.-T.; Lotan, R.; Mercurio, A. M.; Monsigny, M.; Pillai, S.; Poirer, F.; Raz, A.; Rigby, P. W. J.; Rini, J. M.; Wang, J. L. *Cell* **2017**, *76* (4), 597.
(96) Barondess, S. H.; Cooper, D. N. W.; Gitts, M. A.; Lefflerm, H. *J. Biol. Chem*. **1994**, *269* (33), 20807.
(97) Rini, J. M.; Lobsanov, Y. D. *Curr. Opin. Struc. Biol.* **1999**, *9*, 578.
(98) Kasai,K.I., Hirabayashi, J., *J. Biochem*. **1996**, *19* (1), 1.
(99) Kaltner, H.; Gabius, H.-J. *Histol. Histopathol.* **2012**, *27* (4), 397.
(100) Gabius, H. *Eur. J. Biochem.* **1997**, *243*, 543.
(101) Ahmad, N.; Gabius, H.-J.; André, S.; Kaltner, H.; Sabesan, S.; Roy, R.; Liu, B.; Macaluso, F.; Brewer, C. F. *J. Biol. Chem.* **2004**, *279* (12), 10841.
(102) Nieminen, J.; Kuno, A.; Hirabayashi, J.; Sato, S. *J. Biol. Chem.* **2007**, *282* (2), 1374.
(103) Sato, S. *Trends Glycosci. Glycotechnol.* **2002**, *14* (79), 285.
(104) Sato, S.; Nieminen, J. *Glycoconjugate J.* **2004**, 19, 583.
(105) Kilpatrick, D. C. *Biochim. Biophys. Acta* **2002**, *1572*, 187.
(106) Vasta, G. R. *Nat. Rev. Microbiol.* **2009**, *7* (6), 424.
(107) Arnoys, E. J.; Arnoys, E. J.; Wang, J. L. *Acta Histochem* **2014**, 109, 89.
(108) Fukumori, T.; Takenaka, Y.; Yoshii, T.; Kim, H.-R. C.; Hogan, V.; Inohara, H.; Kagawa, S.; Raz, A. *Cancer Res.* **2003**, *63* (23), 8302 LP.
(109) Hsu, D. K.; Yang, R.-Y.; Pan, Z.; Yu, L.; Salomon, D. R.; Fung-Leung, W.-P.; Liu, F.-T. *Am. J. Pathol.* **2000**, *156* (3), 1073.
(110) Fernández, G. C.; Ilarregui, J. M.; Rubel, C. J.; Toscano, M. A.; Gómez, S. A.; Beigier Bompadre, M.; Isturiz, M. A.; Rabinovich, G. A.; Palermo, M. S. *Glycobiology* **2005**, *15* (5), 519.
(111) *Galectins*; Klyosov, A. A., Witczak, Z. J., Platt, D., Eds.; WILEY.
(112) Takemoto, D. K.; Skehel, J. J.; Wiley, D. C. *Virology* **1996**, *458* (217), 452.
(113) Gambaryan, A. S.; Tuzikov, A. B.; Piskarev, V. E.; Yamnikova, S. S.; Lvov, D. K.; Robertson, J. S.; Bovin, N. V; Matrosovich, M. N. *Virology* **1997**, *232*.
(114) Spaltenstein, A.; Whitesides, G. M. *J. Am. Chem. Soc.* **1991**, *113* (2), 686.
(115) Lees, W. J.; Spaltenstein, A.; Kingery-Wood, J. E.; Whitesides, G. M. *J Med Chem* **1994**, *37*.
(116) Sabesan, S.; Duus, J.; Domaille, P.; Kelm, S.; Paulson, J. C. *J. Am. Chem. Soc.* **1991**, *113* (15), 5865.
(117) Joziasse, D. H.; Oriol, R. *BBA - Mol. Basis Dis.* **1999**, 1455 (2-3), 403.
(118) Galili, U. *Immunol. Cell Biol.* **2005**, *83* (6), 674.

(119) Chen, G.; Qian, H.; Starzl, T.; Sun, H.; Garcia, B.; Wang, X.; Wise, Y.; Liu, Y.; Xiang, Y.; Copeman, L.; Liu, W.; Jevnikar, A.; Wall, W.; Cooper, D. K. C.; Murase, N.; Dai, Y.; Wang, W.; Xiong, Y.; White, D. J.; Zhong, R. *Nat. Med.* **2005**, *11* (12), 1295.
(120) Fernández-Tejada, A.; Cañada, F. J.; Jiménez-Barbero, J. *Chem. Eur. J.* **2015**, *21* (30), 10616.
(121) Tangutur, A. D.; Kumar, D.; Krishna, K. V.; Kantevari, S. *Curr. Top. Med. Chem.* **2017**, 2523.
(122) Honore, S.; Pasquier, E.; Braguer, D. *Cell. Mol. Life Sci.* **2005**, *62* (24), 3039.
(123) Jordan, M. *Curr. Med. Chem. Agents* **2012**, *2* (1), 1.
(124) KIRSCHNER, M. W.; MITCHISON, T. I. M. *Nature* **1986**, *324* (6098), 621.
(125) Nogales, E. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30* (1), 397.
(126) Wilson, L.; Jordan, M. A. *Chem. Biol.* **1995**, *2* (9), 569.
(127) Khodiyar, V. K.; Maltais, L. J.; Sneddon, K. M. B.; Smith, J. R.; Shimoyama, M.; Cabral, F.; Dumontet, C.; Dutcher, S. K.; Harvey, R. J.; Lafanechère, L.; Murray, J. M.; Nogales, E.; Piquemal, D.; Stanchi, F.; Povey, S.; Lovering, R. C. *Genomics* **2007**, *90* (2), 285.
(128) Kavallaris M., Don S., Verrills N.M. (2008) *Microtubule-Associated Proteins and Microtubule-Interacting Proteins*. In: Fojo T. (eds) The Role of Microtubules in Cell Biology, Neurobiology, and Oncology. *Cancer Drug Discovery and Development*. Humana Press.
(129) Waterman-Storer, C. M.; Salmon, E. D. *Curr. Biol.* **2017**, *7* (6), R369.
(130) Prota, A. E.; Bargsten, K.; Northcote, P. T.; Marsh, M.; Altmann, K. H.; Miller, J. H.; Díaz, J. F.; Steinmetz, M. O. *Angew. Chem. Int. Ed.* **2014**, *53* (6), 1621.
(131) Altmann, K.-H.; Höfle, G.; Müller, R.; Mulzer, J.; Prantz, K. In *Progress in the Chemistry of Organic Natural Products*; Springer, Vienna, 2009; p 260.
(132) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737.
(133) Felsenfeld, G.; Davies, D. R.; Rich, A. *J. Am. Chem. Soc.* **1957**, *79* (8), 2023.
(134) Jain, A.; Wang, G.; Vasquez, K. M. *Biochimie* **2008**, *90* (8), 1117.
(135) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2006**, *34* (19), 5402.
(136) Ali, A.; Bhattacharya, S. *Bioorg. Med. Chem.* **2014**, *22* (16), 4506.
(137) Baraldi, P.; Bovero, A.; Fruttarolo, F.; Preti, D.; Aghazadeh Tabrizi, M.; Giovanna Pavani, M.; Romagnoli, R. *Med. Res. Rev.* **2004** 24 (4), 475.
(138) Escudé, C.; Sun, J.-S. In *DNA Binders and Related Subjects. Topics in Current Chemistry*; Waring, M. J., Chaires, J. B., Eds.; Springer Berlin Heidelberg, 2005; pp 109–148.
(139) Pabo, C. O.; Sauer, R. T. *Annu. Rev. Biochem.* **1992**, *61* (1), 1053.
(140) Latchman, D. S. *Int. J. Biochem. Cell Biol.* **1997**, *29* (12), 1305.
(141) Ramji, D. P.; Foka, P. *Biochem. J.* **2002**, *365* (Pt 3), 561.
(142) Lekstrom-Himes, J.; Xanthopoulos, K. G. *J. Biol. Chem.* **1998**, *273* (44), 28545.
(143) Kerppola, T.; Curran, T. *Nature* **1995**, *373* (6511), 199.
(144) Vazquez, M. E.; Caamano, A. M.; Mascarenas, J. L. *Chem. Soc. Rev.* **2003**, *32* (6), 338.
(145) Sánchez, M. I.; Vázquez, O.; Martínez-Costas, J.; Vázquez, M. E.; Mascareñas, J. L. *Chem. Sci.* **2012**, *3* (7), 2383.
(146) Rodríguez, J.; Mosquera, J.; Couceiro, J. R.; Vázquez, M. E.; Mascareñas, J. L. *Chem. Sci.* **2015**, *6* (8), 4767.

Chapter 2

Objectives

From the formation side, the key objective of this Thesis is to acquire knowledge on the application of NMR methods to study molecular recognition events.

In the scientific context, the general objective of the thesis is to apply ligand- and receptor-based NMR methods to study the interaction of a variety of ligands of diverse chemical nature with receptors of biological and/or biomedical interest. Within the explored ligands, glycans have been specially employed. Thus, one specific objective is to advance in the understanding of the structural features that govern glycan interactions, including the derivation of epitopes.

In particular, in the glycoscience field, the specific objectives have aimed to:

-Define the molecular recognition features of a synthetic Tn-antigen mimetic.

-Demonstrate the multivalency of a bifunctional ligand towards different targeted proteins.

-Explore the binding of monoclonal antibodies towards carbohydrate-based antigens.

-Provide insight into binding mode and specificity of different Hemagglutinin strains towards sialic acid epitopes.

In the general scientific arena, the specific objectives have aimed to:

-Employ protein-based NMR methods to characterize the intrinsic structural propensity of a DNA-binding protein motif.

-Characterize the bound conformation of a novel epothilone derivative towards tubulin, both in its dimeric state and when forming microtubules.

-Assess the possibilities of different ligand-based NMR methods to screen dynamic combinatorial libraries in the presence of a given receptor.

-Understand the specificity and the interaction features of novel DNA minor groove binders towards a consensus DNA sequence.

Chapter 3

NMR applications to carbohydrate recognition

3.1. Introduction

In the era of the omics, glycomics has become important given the ubiquitous presence of glycans in biological processes. The chemical diversity of carbohydrates is rather high. Individual monosaccharides associate in a variety of manners to form polysaccharides displaying multiple connections or oligosaccharides that are decorated with multiple pendant groups, or glycosylated lipids, peptides or proteins. All these variations permit to encode biological information that is transmitted in a rather complex manner. This translation process is denominated sugar code, which remains largely unexplored. Very probably, the 3D structures of the glycoconjugates and their dynamic features are at the heart of triggering the corresponding biological event, though a cascade of recognition events.

Carbohydrate recognition by receptors is mediated by weak interactions (see introduction) and often displays weak affinities. However, in nature, carbohydrates are usually found in branched and multivalent scaffolds (i.e. multi branched sialylated glycans in the human upper respiratory tracts) that permit to overcome the limitation of weak affinity by the cooperative binding of several carbohydrates to the receptors. Therefore, carbohydrate multivalency is a key principle in nature to achieve strong, yet reversible interactions (i.e. virus infection).[1] In addition, carbohydrate multivalency has to take into account also in carbohydrate based drug design. The current therapeutic approach is to employ monovalent drugs and consequently, high doses are required to be effective. As a result, the achievement of multivalent scaffold architectures has huge potential in the therapeutic action of carbohydrate-based drugs.[1–3]

The numerous roles of glycoconjugates in life processes are mediated by binding proteins that decode the information content in the glycome through their recognition of glycans. Therefore, the understanding of the basis of the molecular recognition between glycoconjugates and their different receptors is mandatory for the rational design of proper drugs. In this context, NMR arises as a valuable tool for gathering essential information and providing key knowledge on molecular recognition and a 3D view of the ligand-receptor complex.

In the following sections of this chapter, I will describe different examples of the application of different NMR methods, in many occasions assisted by molecular modelling techniques, to analyse the interaction of glycans by different receptors, lectins and antibodies.

3.2. Structural insights in the recognition of the Tn antigen by lectins

In the late 50s, an antigen similar to already known T antigen was first observed by Moureau *et al*. They dubbed it as "T antigen nouvelle" (Tn antigen, figure 1).[4] The simple chemical structure of the Tn antigen consists of a GalNAc residue α-linked to the hydroxyl group of a Ser or Thr amino acid residue. A decade later, the presence of this Tn antigen in tumor cells was described, opening its study as a "tumor associated carbohydrate antigen" (TACA). The Tn antigen is found in more than the 90% of the tumor cells,[5] since it is overexpressed in glycoproteins such as O-glycosylated mucins in the cell membrane. In fact, it could be considered as a cancer biomarker due to this atypically high abundance in tumors. Despite its simplicity, the Tn antigen plays a key role in numerous human disorders starting from the Tn syndrome,[4,6–8] cancer,[5,9–12] HIV infection[13] or IgA nephropathies.[14] For instance, the aberrant glycosylation of membrane proteins produce an overexpression of the Tn antigen[15] in more than 90% of breast tumors. Moreover, the presence of the Tn antigen is directly correlated with metastatic potential and poor prognosis.[15–18]

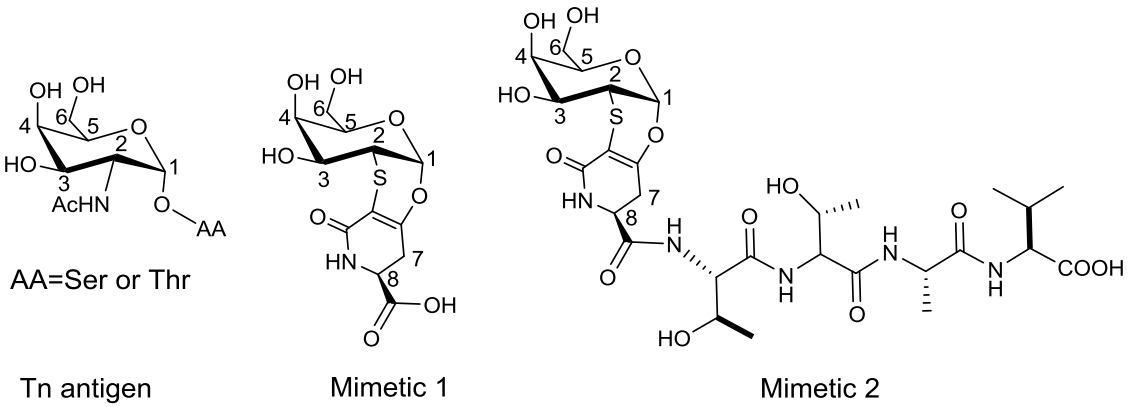
It has been reported that the Tn antigen is bound by the C-type macrophage galactose binding lectin (MGL), deactivating the immune response and thus enabling the tumor to escape from inmunosurveillance triggering metastasis and tumor progression.[18–20] Thus, the implication of the Tn antigen in cancer and other serious pathologies has been studied during the last years and consequently considered as a potential therapeutic target. Many efforts are currently underway worldwide to develop a carbohydrate-based vaccine against the Tn antigen. However, the appropriate epitope with similar binding properties has not yet been discovered.

The synthesis of Tn mimetics and Tn antigen epitopes is one of the open ways towards the development of antitumor vaccines. Because of its carbohydrate nature and its relative small size as an antigen, the vaccination process has to overcome the key problems related to the presentation of the epitopes to the immune system and the presence of glycosidases.[21–24] Thus, new methods have to be developed to enhance the presentation of the epitope and then, the concomitant immune response. The success of the synthesis of Tn epitope mimetics resides in the recognition of these analogues by anti-Tn antibodies with the same efficiency as the "natural" Tn antigen.

The work exposed in the following pages present the study of the structural and conformational properties of new Tn antigen mimetics (mimetic **1** and mimetic **2**, scheme 1) developed by the group of Prof. Cristina Nativi at University of Florence[25] as potential Tn

vaccination substitutes. The recognition properties have been compared to those of a Tn simplified model (α-OMeGalNAc) and its related monosaccharide, galactose.

The mimetic **1** consists of a tricyclic structure that can be seen as a conformationally constrained mimetic that maintains the key structural features of Tn antigen. In contrast, mimetic **2** is the glycopeptide extension of **1** mimicking the glycosylated N-terminus STTAV sequence of mucin CD43.[26,27] The carboxylic moiety at the end of both structures provides these two molecules with the possibility of performing different chemical modifications at these positions to achieve the proper presentation into multivalent architecure (i.e. nanoparticle functionalization, glycodendrimers, virus-like particles…) for the development of an efficient and effective vaccine.



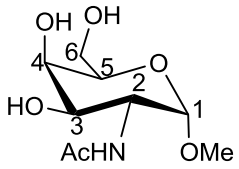
Scheme 1. Scheme of the structures of Tn antigen and mimetics 1 and 2.

From the NMR perspective, the study of the binding properties was addressed from the ligand point of view. The major aim of the study was to obtain its contact epitope information to different Gal/GalNAc-binding lectins. Thus, these mimetics were tested with three lectins from diverse sources; plants (from legume Erythrina Cristagalli Lectin, ECL),[28] mammals (C-type lectin from human, Macrophage Galactose Lectin, MGL),[15] and invertebrates (from snail, Helix Pomatia Agglutinin, HPA)[29] origin.

3.2.1. The plant lectin: *Erythrina Cristagalli* (ECL)

The *Erythrina cristagalli* lectin (ECL) is a galactose (Gal) specific legume lectin isolated from the cockspur coral tree.[28,30] The ECL has been studied since the early 80s, but it was in 2004 when ECL was thoroughly characterized and its affinity towards galactosides was evidenced. Remarkably, ECL displays better affinity towards GalNAc versus Gal.[28] ECL is a glycoprotein of 56.8 kDa consisting in two subunits of similar molecular weight and exhibits hemagglutinating

activity and mitogenic effect over T lymphocytes in in-vitro tests.[31,32] However, its biological functions in the plant itself are still unknown.



α-OMeGalNAc

Analogue **3**

Scheme 2. Scheme of the Tn antigen simple model, α-OMeGalNAc (analogue 3).

First, the STD-NMR-based study of ECL in the presence of the Tn antigen simple model, α-OMeGalNAc (**3**) was successfully performed as a positive control of the binding activity (figure 1).

Then, the STD-NMR of mimetic **1** in the presence of ECL (figure 2) was carried out, clearly showing its recognition. Basically, all protons of mimetic **1** received saturation from the protein. The analysis of the STD spectra of both molecules (**1** and **3**) in the presence of ECL (table 1), indicated that H1 and H2 at the galactose pyranose ring always received the highest saturation from the protein. Some differences between **1** and **3** were observed for protons H3 and H4.

As mentioned in the introduction, a major efficiency in the saturation is translated into a closer distance between the observed nucleus and the protein protons at the binding site. In our particular case, the observed STD intensities indicate that H1 and H2 are much closer to the binding site than H5 and H6.

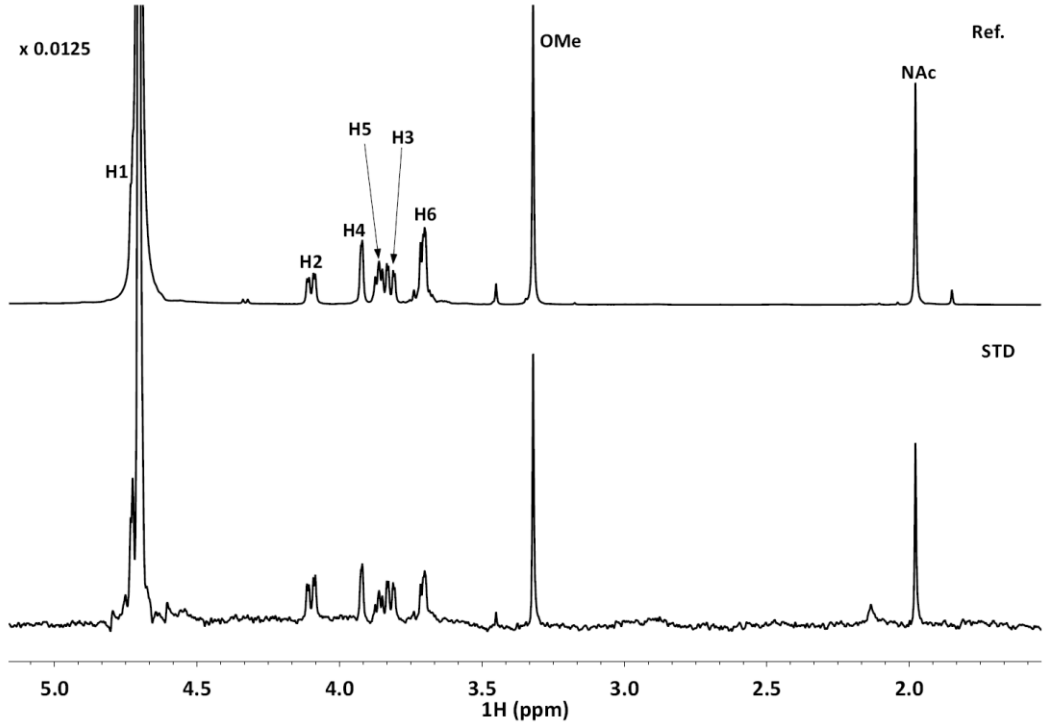


Figure 1. STD (below) and off-resonance (on top) NMR spectra (500 MHz) of a ECL:α-OMeGalNAc mixture (1:100 molar ratio) with 40 μM ECL. The temperature was 298 K and a 2 seconds saturation time at the aromatic region (δ 7 ppm) was employed. A spin lock filter was used to minimize the background of the protein protons.

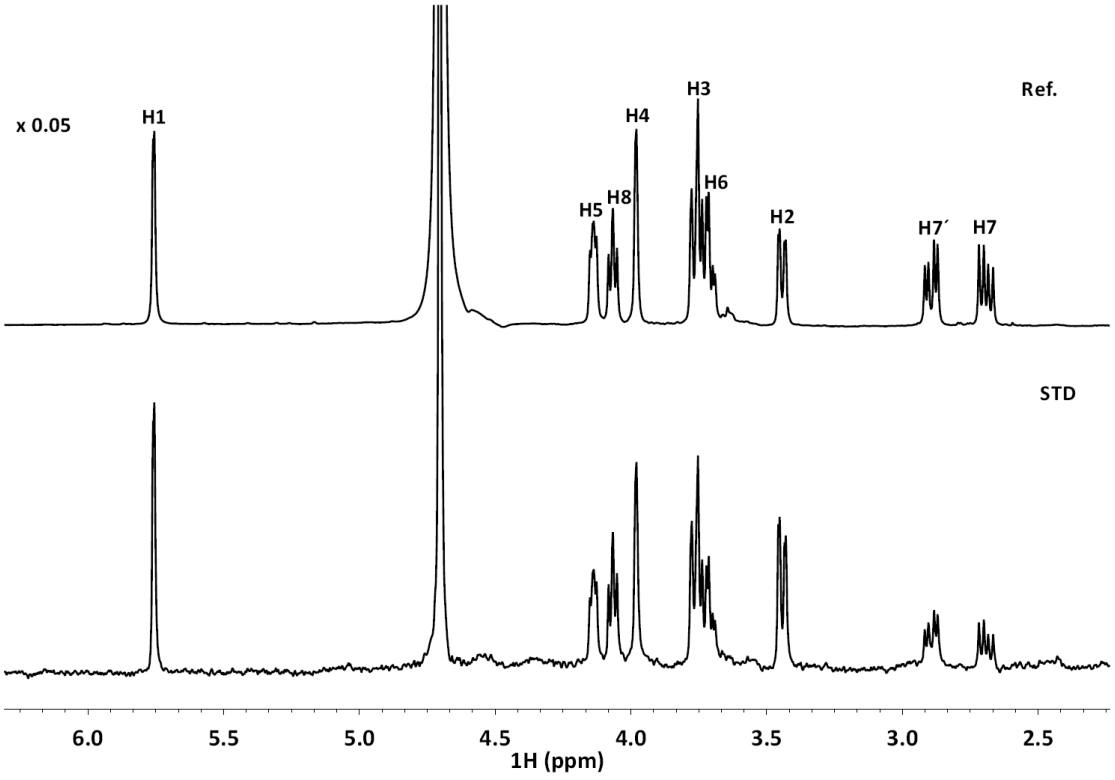


Figure 2. STD (below) and off-resonance (on top) NMR spectra (500 MHz) of a ECL:1 mixture (1:100 molar ratio) with 40 μM ECL. The temperature was 298 K and a 2 seconds saturation time at the aromatic region (δ 7 ppm) was employed. A spin lock filter was used to minimize the background of the protein protons.

	Absolute STD intensity (%)			Relative STD intensity (%)		
	Mimetic **1**	Mimetic **2**	α–OMeGalNAc	Mimetic **1**	Mimetic **2**	α–OMeGalNAc
H1	7.2	4.2	1.7	90	45	98
H2	8.0	7.9	1.7	100	85	100
H3	5.3	5.1	1.3	66	55	78
H4	5.3	3.6	1.4	66	39	83
H5	4.7	4.3	0.9	59	46	51
H6	3.8	3.2	0.8	48	35	49
H7a/H7b	2.6 and 3.3	2.6	-	33 and 41	28	-
H8	5.7	9.3	-	72	100	-
NAc	-		1.0	-	-	60
OMe	-		1.1	-	-	66

Table 1. STD intensities from the STD spectra shown in figures 2, 3 and 6.

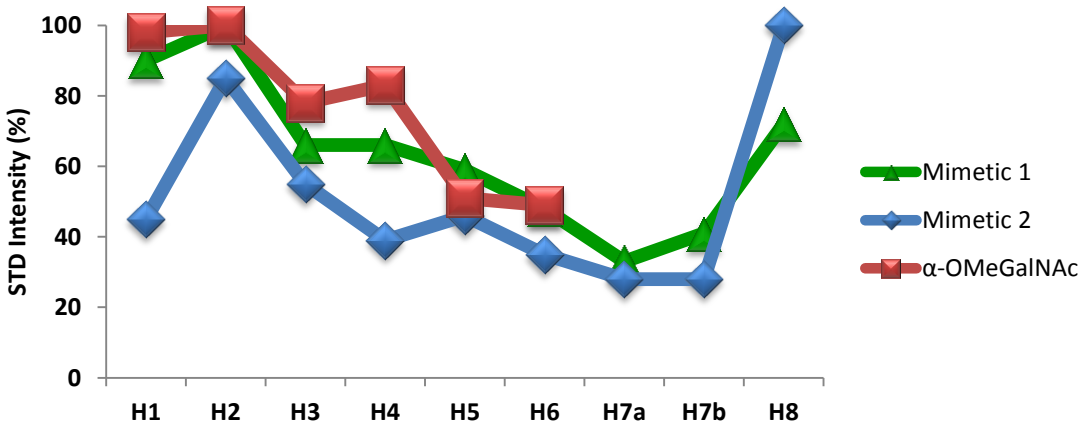


Figure 3. Observed STD intensities for the different protons of the model compound and the two mimetics when interacting with ECL. The data mimetic 1 (in green), α–OMeGalNAc (in red) and mimetic 2 (in blue).

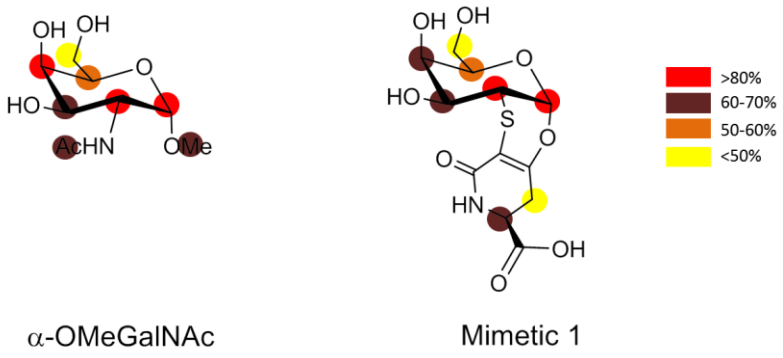


Figure 4. Epitope mapping deduced by STD NMR for the binding of α–OMeGalNAc and mimetic 1 to ECL.

Competition studies were also performed to further elucidate if both molecules share the same binding site. The observation of competition between the two molecules would indicate

that they bind at same binding site whereas the absence of competition will suggest that they bind in different places of the protein (unless allosteric behavior is suspected). To address competition by NMR, ligand displacement experiments were performed starting with a sample of α–OMeGalNAc as reference compound in presence of the lectin. The experiment was carried out by monitoring the evolution of the STD signal intensities of selected and isolated protons (in terms of chemical shift) of the reference compound upon increasing additions of the competitor. Figure 5 shows up the evolution of the signals at increasing amounts of competitor (mimetic **1**). Due to the typical signal overlapping of carbohydrates, the selection of the signal to be monitored was not trivial. Fortunately, H4 from mimetic **1** and O-methyl group of α-OMeGalNAc appeared clearly isolated and then were monitored to follow the competition process.

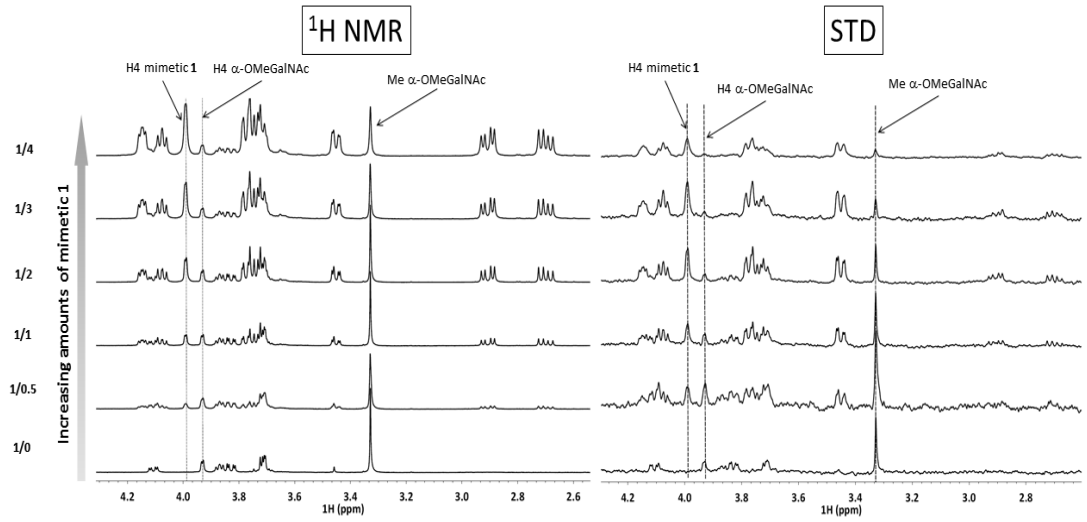


Figure 5. Competition STD-NMR between α–OMeGalNAc and 1 with ECL (500 MHz, 298 K). On the left handside, off-resonance upon increasing amounts of 1 up to molar ratio 1:4 α-OMeGalNAc:1. On the right, the corresponding STD spectra of each ratio. The concentration of ECL was 10 μM.

The competition study revealed that mimetic **1** is able to displace α-OMeGalNAc from the binding site, thus demonstrating that both molecules compete for the same binding pocket of the protein. When both compounds are in 1:0.5 α-OMeGalNAc:**1** ratio, H4 signal from mimetic **1** in the STD is similar than the same H4 signal from α-OMeGalNAc. When higher molar ratios of mimetic **1** were employed, the H4 and methyl signals of analogue **1** were significantly reduced, thus indicating that the analogue is removed from the binding site.

As next step, the behavior of glycopeptide Tn mimetic **2** was studied. In this case, the tricyclic compound is connected to a peptide fragment derived from the N-terminal of CD43 glycoproteins, trying to mimic the Tn antigen presentation in mucins. Thus, the STD NMR spectrum of mimetic **2** in the presence of ECL (molar ratio 1:100) was obtained (figure 6). It can be observed that only the protons from the sugar part provide STD intensities, while the signals corresponding to the peptide fragment barely show up. The epitope mapping for **2** is plotted in figure 3 in comparison with that deduced for mimetic **1** and α-OMeGalNAc. Even though some differences were observed, especially for H4, the epitope profile follows a similar pattern for both mimetics, **1** and **2**, with H2 receiving the highest saturation from ECL. On the other hand, the peptide part of the mimetic **2** gives very weak STD effect, only observable for the Thr moiety close to the pyranose ring. Only ca. 20% of STD relative intensity was measured (100% for H2, table 1) for Hα and Hβ protons, with 6% for the methyl groups. All these results indicate that the peptide has marginal contacts with the lectin, being the sugar part the key interacting epitope.

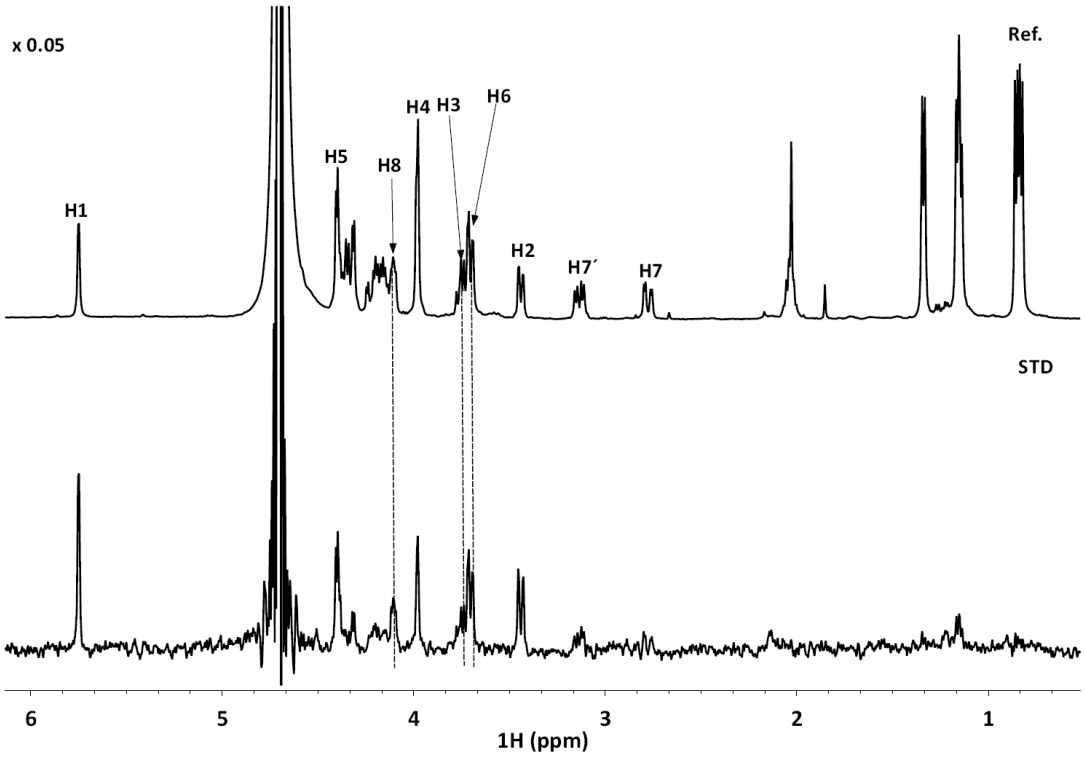
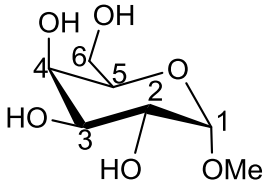


Figure 6. The interaction of 2 with ECL. The STD-NMR spectrum (500 MHz, 298 K) obtained for a 100:1 mixture of 2:ECL is shown at the bottom, while the off-resonance spectrum is shown at the top. The concentration of the lectin was 40 μM. A spin lock filter was used to minimize the background of the protein signals, while the on-resonance frequency for protein saturation was set at the aromatic region (δ 7ppm).

3.2.2. The human lectin: *Macrophage Galactose Lectin* (MGL)

The *Macrophage Galactose Lectin* (MGL, CD301) is a transmembrane receptor present in cells of the immune system,[33] with a C-type lectin domain in its extracellular part. The C-type lectin family is defined by its calcium dependence for binding and accommodate a calcium ion in their structure which is essential to shape the binding site and directly interacts with the ligand.[34] MGL specifically binds glycans with terminal GalNAc residues and has been associated with the immune response to the pathologies produce by the Tn antigen.[15] Previous works have reported that MGL recognize the Tn antigen located at the surface of the glycosylated mucins present in numerous carcinomas such as breast or colon and it can be involved in the modulation of the immune response towards some tumor cells.[18]

Applying the same methodology as described above, a positive control STD experiment (figure 7) with α-OMeGal (**4**, scheme 3), showed that H4 provided the strongest STD intensity, while the other protons present in the pyranose ring are weaker as previously reported.[35,36]



α-OMeGal

Analogue **4**

Scheme 3. Structure of the α-OMeGal (analogue 4).

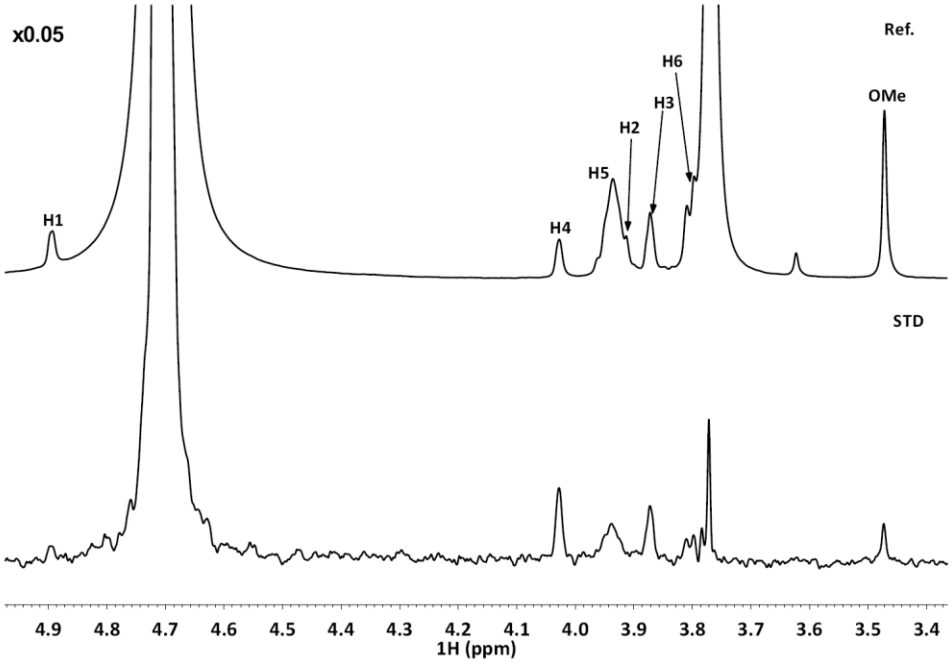


Figure 7. The interaction of α-OMeGal with MGL. The STD-NMR spectrum (500 MHz, 308 K) is given below, while the off-resonance spectrum is given on the top. A 100:1 α-OMeGal:MGL molar ratio was employed. The measured relative STD intensities were: H1 21%, H2+H3 51%, H4 100%, H5 20%, H6 14%, OMe 13%.

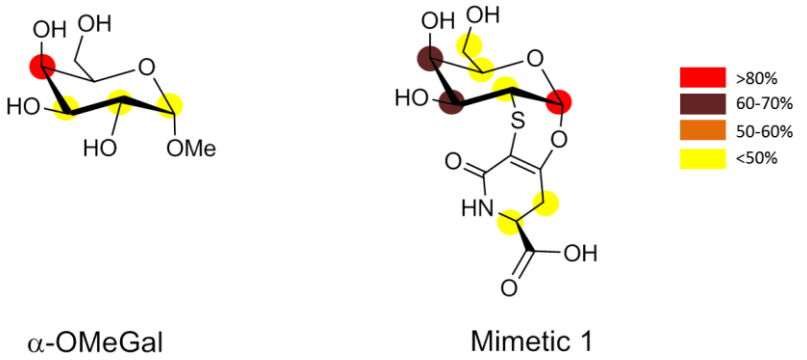


Figure 8. Epitope mapping results for the binding of α-OMeGal and 1 to MGL.

Next, the STD-NMR study of mimetic **1** in the presence of MGL was performed. Interestingly, the deduced STD-derived epitope of **1** to MGL was different to that found for α-OMeGal epitope (figure 8). For **1**, H1 showed the highest STD value, providing an epitope closer to that already described for α-OMeGalNAc[35] where protons H1 and H2 receiving the highest saturation. STD-NMR-based competition experiments were then carried out. Initial competition studies showed that mimetic **1** was not able to displace α-OMeGalNAc from the binding site of MGL. Thus, STD-NMR competition experiments (figure 9) were carried out using α-OMeGal, although it is a weaker ligand than α-OMeGalNAc for MGL.[35] The STD intensities of

H4 of α-OMeGal and mimetic **1** were followed upon the step-wise addition of mimetic **1** to the mixture solution (figure 9). A sevenfold excess of mimetic **1** with respect to α-OMeGal was necessary to reach a significant competition (H4 signals form mimetic **1** and α-OMeGal showed then similar intensities and the methyl signal experienced a clear reduction, figure 9). These results indicate that mimetic **1** binds to MGL even more weakly than α-OMeGal does, probably over the millimolar range.

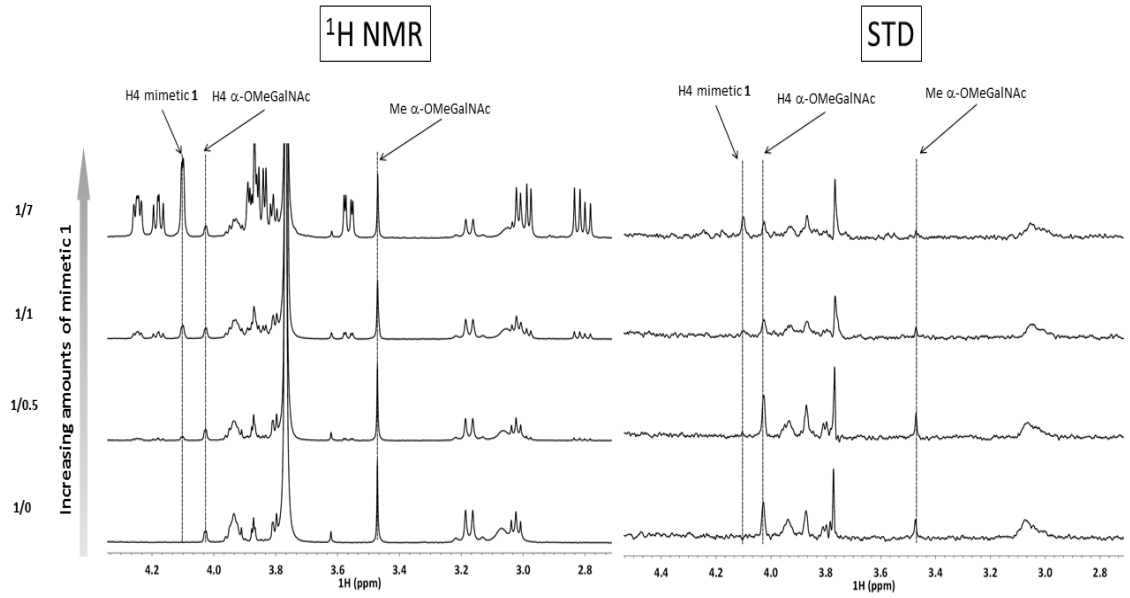


Figure 9. Competition STD-NMR between α–OMeGal and 1 with MGL (500 MHz, 308 K). On the left-hand side, the off-resonance spectra upon increasing amounts of 1 up to 1:7 molar ratio α-OMeGalNAc:1. On the right-hand side, the corresponding STD spectra of each ratio α-OMeGalNAc:1. The concentration of MGL was 10μM.

3.2.3. The invertebrate lectin: *Helix Pomatia Agglutinin* (HPA)

Lectins extracted from snails are known to recognize blood group antigens. In the present context, the recognition of the Tn antigen by the *Helix Pomatia Agglutinin* (HPA), a lectin form the edible snail, was characterized in 1969 by Prokop *et al*[37] being a milestone in TACAs research, due to its relevance in breast tumors and carcinomas.[38] HPA specifically binds to terminal α-GalNAc residues,[29] and therefore it was included as an additional model lectin to test the binding properties of the Tn antigen mimetics.

The STD spectrum of HPA in the presence of mimetic **1** (figure 10) showed clear saturation transfer signals. The highest saturation was shown by H1 and H2, whereas H7 displayed no intensity. The STD intensities were then compared to those obtained for the positive control, α-OMeGalNAc (figure 11).

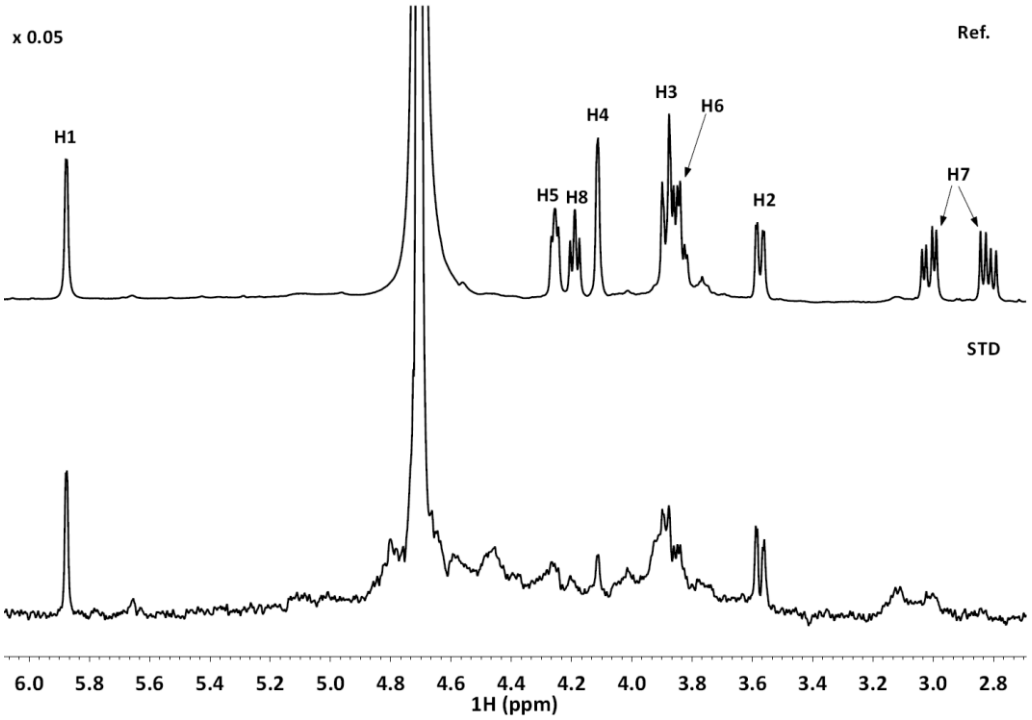


Figure 10. The interaction of 1 with HPA. The STD spectrum (500 MHz, 308 K) obtained for a 100:1 mixture of 1:HPA is shown at the bottom, while the off-resonance spectrum is shown at the top. The concentration of the lectin was 40 μM. A spin lock filter was used to minimize the background of the protein signals, while the on-resonance frequency for protein saturation was set at the aromatic region (δ 7ppm).

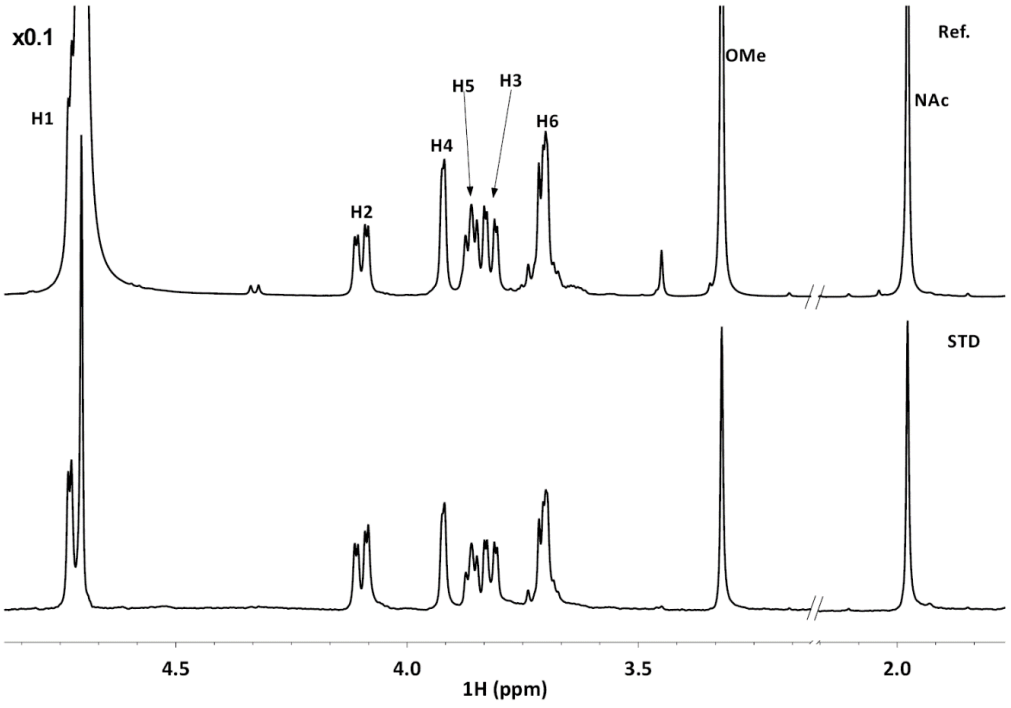


Figure 11. The interaction of the model compound, α-OMeGalNAc, with HPA. The STD spectrum (500 MHz, 308 K) obtained for a 100:1 mixture of α-OMeGalNAc:HPA is shown at the bottom, while the off-reference spectrum is shown at the top. The concentration of the lectin was 40 μM. A spin lock filter was used to minimize the background of the protein signals, while the on-resonance frequency for protein saturation was set at the aromatic region (δ 7ppm).

	Absolute STD intensity (%)		Relative STD intensity (%)	
	Mimetic **1**	α–OMeGalNAc	Mimetic **1**	α–OMeGalNAc
H1	5.1	8.5	100	76
H2	5.1	11.1	100	100
H3	2.5	7.6	49	68
H4	1.3	6.9	26	62
H5	n.d.	6.8	n.d.	61
H6	2.3	5.6	45	50
H7	n.d.	-	n.d.	-
H8	n.d.	-	n.d.	-
Ac	-	5.6	-	50
OMe	-	4.1	-	37

Table 2. STD intensities measured for the interaction of mimetic 1 and α-OMeGalNAc with HPA.

The epitope mapping analysis of the STD spectra obtained permitted deducing that H1 and H2 of mimetic **1** received the highest saturation, while H3, H4 and H6 showed weaker intensities. This fact suggests that the recognition of mimetic **1** mainly involves the carbohydrate structure, while the bicyclic extension is not involved in the molecular recognition event. However, for α-OMeGalNAc, although the major saturation transfer was observed at H2 signal, the whole pyranose ring is recognized by HPA. From a global perspective, the profile of the STD intensities for both molecules (figure 12) is fairly similar, suggesting that both molecules display a similar binding mode to HPA.

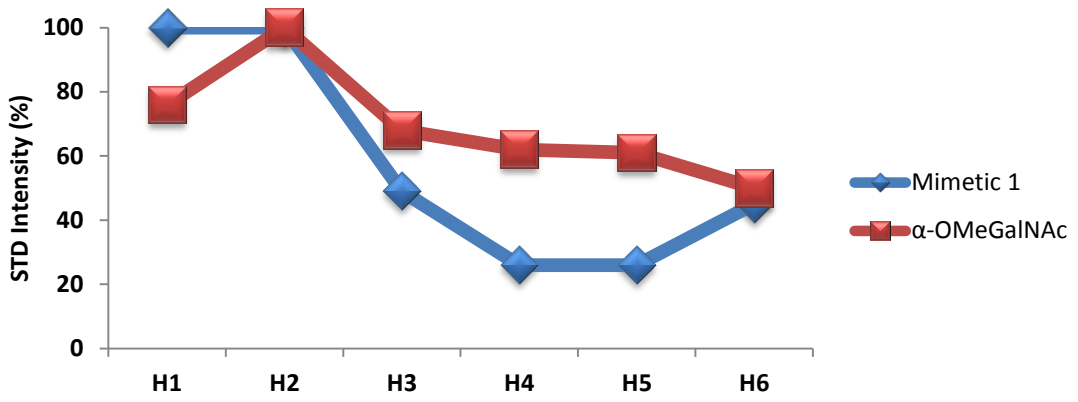


Figure 12. The profile of the STD intensities for each proton of mimetic 1 (in black) and α-OMeGalNAc (in grey) for the interaction with HPA.

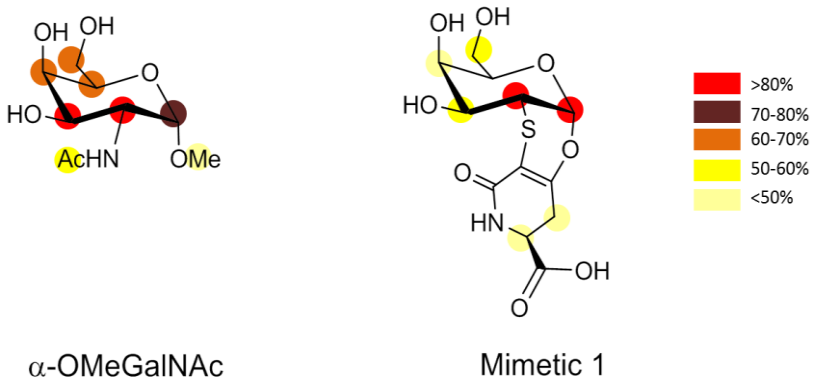


Figure 13. STD-based epitope mapping for the binding of ligands α-OMeGalNAc and mimetic 1 to HPA.

3.2.4. Conclusions

Therefore, the overall picture of the obtained results for the interaction of **1** with the three lectins can be concluded that this mimetic behaves as a reasonable mimetic of the Tn antigen. This molecule presents a similar binding mode to that observed for the Tn model (α-OMeGalNAc), while the rigid scaffold mimetic **1** does not impede the binding. Nevertheless, the relative affinities vary among the three tested lectins. The STD-NMR-based competition experiments revealed that mimetic **1** is able to remove α-OMeGalNAc from the binding site of ECL, indicating a similar or even higher affinity. Interestingly, the data obtained for glycopeptide mimetic **2** have also indicated that ECL recognizes its sugar moiety with a similar epitope to that deduced for the recognition of mimetic **1** and α-OMeGalNAc. Additionally, the analysis of the NMR data has also shown that **1** is also a fair mimic of the Tn model in the interaction with HPA lectin. However, for MGL, much higher concentrations of **1** were needed to partially remove α-OMeGal from the lectin binding site. Therefore, the binding affinity is now rather weak. This fact indicates that the bicyclic moiety strongly interferes with the binding site amino acids. Previous molecular modelling data have shown that the acetamide moiety of Tn directly participates in the interaction with MGL.[35] Therefore, it is highly likely that the chemical modification present in **1** at this moiety strongly affects MGL binding.

All these data permit to consider the tricyclic mimetic **1** as a rigid but selective mimic of the Tn antigen, always depending on the chemical architecture of the receptor binding site.

3.2.5. Experimental

MGL lectin was provided (Prof. H.J. Gabius, LMU, Munich, Germany) as its extracellular fragment (amino acids 68-316), containing the carbohydrate recognition domain (CRD, amino

acids 187-316). The lectins Erythrina Cristagalli (ECL) and Helix Pomatia (HPA) were purchased from Sigma-Aldrich.

The α-OMeGalNAc as well as the α-OMeGal were commercially available from Sigma-Aldrich, while mimetics **1** and **2** were chemically synthesized by Prof. Nativi's group at the University of Florence.[39]

Samples with ECL and HPA were prepared in deuterated PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4; 1.7 mM KH_2PO_4, pH 7.3), while perdeuterated TRIS buffer (10 mM Tris, 1 mM $CaCl_2$, 75 mM NaCl, pH 7.5) in D_2O was used for MGL.

All NMR experiments were acquired on a Bruker Avance instrument operating at 500 MHz (^1H frequency). For the STD-NMR experiments, the saturation frequency was set either at 7 ppm (aromatic region) or at 0 ppm (aliphatic region). All the STD-NMR spectra displayed were collected with a short spin lock filter to minimize the background of the protein signals.

For ECL, the STD-NMR spectra were acquired using a sample containing 40 μM ECL and 4 mM mimetic **1** (molar ratio 1:100) at 298 K. For HPA, the STD spectra were acquired on a sample containing 40 μM lectin and 4 mM mimetic **1** (molar ratio 1:100) at 298 K and 308 K. Finally, for MGL, the STD-NMR spectra were acquired on a sample containing 10 μM of MGL and 1mM of mimetic **1** (molar ratio 1:100) at 308 K.

STD-NMR-based competition experiments

To a solution of 30 μM of ECL and 3 mM of α-OMeGalNAc, increasing amounts of mimetic **1** were added, until a final concentration of 12 mM of **1** was reached (4-fold excess with respect to α-OMeGalNAc) (table 3). The STD-NMR spectra were recorded in a 500 MHz spectrometer at 298 K.

	ECL (μM)	α-OMeGalNAc (mM)	Mimetic 1 (mM)
Exp. 1	30	3	0
Exp. 2	30	3	1.5
Exp. 3	30	3	3
Exp. 4	30	3	6
Exp. 5	30	3	7.5
Exp. 6	30	3	9
Exp. 7	30	3	12

Table 3. Description and summary of the experiments and concentrations employed for the STD-NMR-bsaed competition assays with ECL.

In the case of MGL, to a solution of 10 μM MGL and 1 mM α-OMeGal, increasing amounts of **1** were added until a final concentration of 8.1 mM of **1** was reached (8-fold excess with respect to α–OMeGal) (table 4). The STD-NMR spectra were recorded in a 500 MHz spectrometer at 308 K.

	MGL (μM)	α-OMeGal (mM)	Mimetic 1 (mM)
Exp. 1	10	1	0
Exp. 2	10	1	0.3
Exp. 3	10	1	1.7
Exp. 4	10	1	8.1

Table 4. Description and summary of the experiments and concentrations employed for the STD-NMR-based competition assays with MGL.

3.3. Bifunctional ligands with one glycan epitope: Towards the simultaneous target of two different proteins

As mentioned above, carbohydrates participate in the early stages of diverse inter and extra cellular recognition events that facilitate the interaction between partners (from small molecules to large entities: cells, bacteria, viruses, etc.) to start or to maintain a chain of subsequent biological processes. As paradigmatic examples, the rolling of leukocytes in blood vessels[40] or recognition of pathogens to initiate defense response can be mentioned.[41]

In this context, the family of the galectins presents diverse functions at the extracellular matrix (ECM) or inside the cell. Galectins are a type of carbohydrate-binding proteins (lectins) that have the ability to recognize β-galactosides.[42] All the galectins contain a well conserved carbohydrate binding domain (CRD) responsible for carbohydrate binding. Many families of galectins have already been described and classified. Some of them show one CRD, while other show two CRDs, usually functioning in a non-cooperative manner. Galectin 3 (*h*gal-3) is the only galectin that contains a tail of repetitive amino acids attached to its single CRD.[43,44] It has been described that the cytoplasmic upregulation of this chimera-type lectin has anti-apoptotic effects,[45–47] whereas nuclear and extracellular *h*gal-3 produces completely opposite pro-apoptotic or even anti-apoptotic activity, depending on the type of cells or its intracellular localization.[48–52] Interestingly, human gal-3 (*h*gal-3) has been described to form oligomeric structures (pentamers) in the extracellular environment.[53]

The presence of *h*gal-3 in the extracellular matrix (ECM) has been correlated with the activity of some matrix metalloproteinases (MMPs) being itself a substrate of MMPs. These enzymes, remove the tail from *h*gal-3, impairing the ability of oligomerization.[53–59] MMPs are a large

class of at least 23 different zinc-dependent enzymes present in the extracellular matrix that are involved in many physiological cellular processes and its dysregulation produces pathologies, tumor growth or metastasis. Of note, galectins can enhance MMPs production and malignancy, related to tumor growth and progression. Thus, the association of *h*gal-3 in ECM to many tumors makes *h*gal-3 a therapeutic target. Moreover, since there are MMPs in the ECM involved in tumor progression, the design of multiple valency ligands to simultaneously target both receptors might be valuable to fight against the pathology. However, MMP inhibitors have failed in clinical trials because of limited solubilities in physiological conditions and low selectivities due to the MMP remarkable property to adapt their binding pocket to inhibitor´s shape. Thus, dual-target-related molecules may be a new approach to overcome the limitation of scarce selectivity of the MMP inhibitors.[60]

As mentioned above, multivalent processes are a common situation in carbohydrate-protein interactions. In many occasions, the ligand (sugar) presentation is multivalent. In fact, in many glycoconjugates present multibranched structures that contains several copies of the same carbohydrate epitope in the same molecule. Therefore, it can simultaneously bind several carbohydrate receptors. On the other hand, the glycan receptors, e.g. lectins can have more than one carbohydrate recognition domain in the same molecule or exist in a multimeric. In all cases, the receptors simultaneously present several carbohydrate binding sites.

The work present here represents the continuation of our previous studies performed in collaboration with the group of Prof. Cristina Nativi at the University of Florence. We have reported on the design of bifunctional linear compounds containing the lactose moiety on one end and a metalloproteinase inhibitor of sulfamide-hydroxamate type[61–63] on the other end. Our results showed the feasibility of this approach, since the synthetic compound was able to simultaneously bind human gal-3 carbohydrate recognition domain (*h*gal-3 CRD) at one end (the glycan epitope) and the catalytic domain of MMP-12 (catMMP-12) at the alternative end (the hydroxamate epitope).[64]

In this work, and as logical extension of our initial findings, we have explored a multivalent strategy. Thus, a water soluble MMP inhibitor that incorporates two glycan branches (figure 14) for an increased valency to enhance the avidity for *h*gal-3 CRD has been employed and its molecular recognition properties by both catMMP-12 and *h*gal-3 CRD has been evaluated. The molecular recognition events have been followed by using different NMR strategies, including those based on ligand observation, as STD-NMR experiments, and also others based on the

observation of the protein signals by means of HSQC-based experiments, using isotopically ^{15}N labeled proteins.

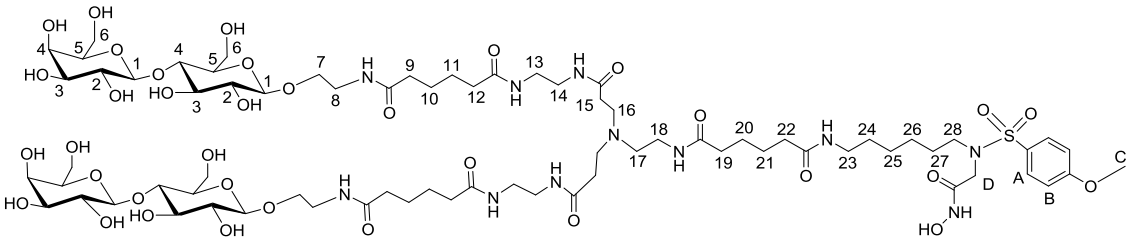


Figure 14. Scheme of compound 3.

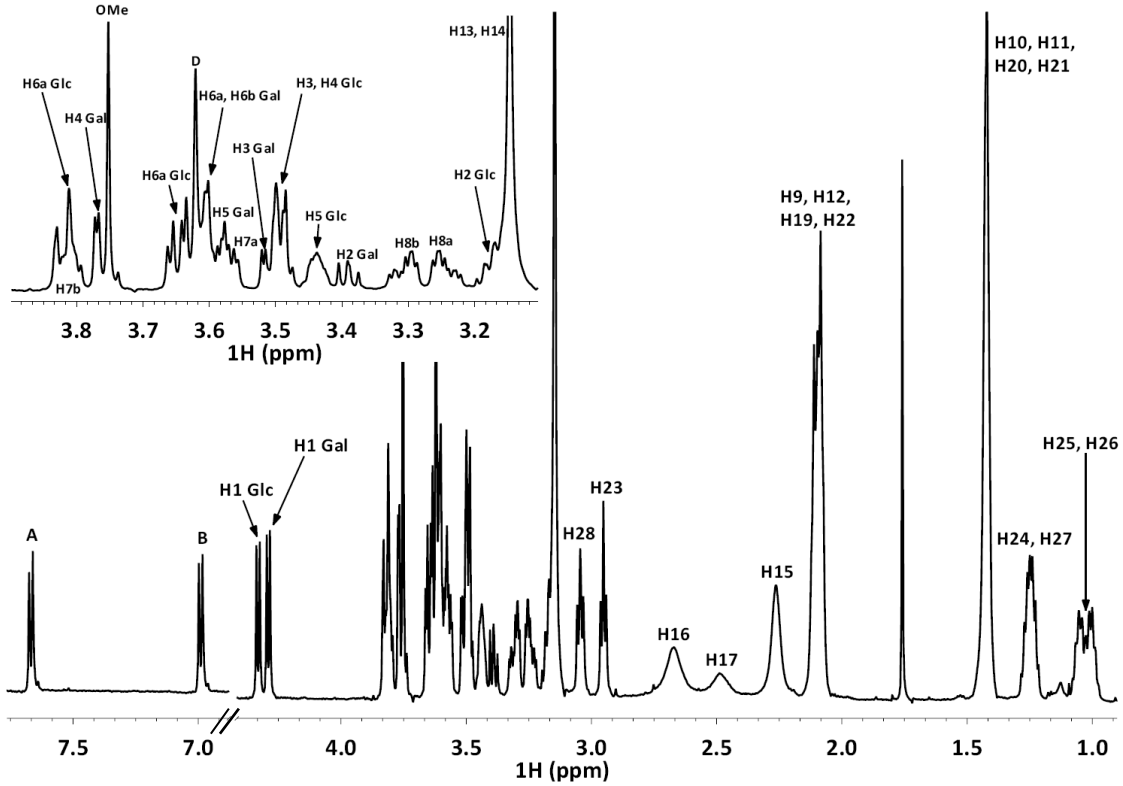


Figure 15. The ^{1}H NMR spectrum of compound 3 at 600 MHz (298 K) with the assignment of different signals.

After assigning the different NMR signals of **3** using standard 2D NMR methods, the interaction with the two proteins was studied independently. The interaction of **3** with catMMP-12 was followed from the protein´s point of view using a chemical shift perturbation analysis of the ^{1}H-^{15}N amide cross peaks measured in ^{1}H-^{15}N HSQC spectra obtained for the catMMP-12 in the absence and presence of the ligand. Additionally, the interaction of **3** with *h*gal-3 CRD was studied both from the ligand´s point of view through STD-NMR and from protein´s point of view using analogous approach to that described above for catMMP-12. Finally, the possible

formation of a ternary complex was studied by monitoring the changes in the shapes of the cross peaks in the ^{1}H-^{15}N HSQC spectra of ^{15}N labeled catMMP-12 bound to **3**, in the absence and in the presence of *h*gal-3 CRD.

3.3.1. The interaction of **3** with *h*gal-3 CRD

The existence of interaction of ligand **3** with *h*gal-3 CRD was first evidenced by the mere observation of its standard 1D ^{1}H NMR spectrum in the presence of the lectin. The existence of significant line broadening effects in the resonance signals of the ligand when the lectin was added to the solution indicated the presence of a binding event. Indeed, since the linewidth of the NMR signals is directly related to transverse relaxation rate ($1/T_2$), the binding of a medium-small molecule (such as **3**) to a large receptor strongly modifies its relaxation properties, a phenomenon that can be monitored at the line width level. In fact, high molecular weight proteins show broad signals whilst low molecular weight ligands show narrow NMR signals. Thus, the observation of broadening effects for the NMR signals of a putative ligand in the presence of the receptor indicates a reduction in the transversal relaxation T_2 that can be directly associated with the formation of a molecular complex.

The line broadening of the ligand´s signals in the NMR spectra acquired for two samples with different ligand/protein ratios (10:1 and 50:1) were compared to that observed in the absence of the receptor (bottom panel of figure 16). As shown in figure 16, at a of 10:1 molar ratio (**3:***h*gal-3 CRD) significant broadening of the NMR signals of the ligand was observed, especially for protons H1 and H4 of the terminal Gal residue, and to a minor extent for the anomeric H1 of Glc, while other protons, such as H8, A or B, were less affected. At a higher **3**:*h*gal-3 CRD molar ratio (50:1), the signal broadening was partially reverted towards that of the free ligand in solution.

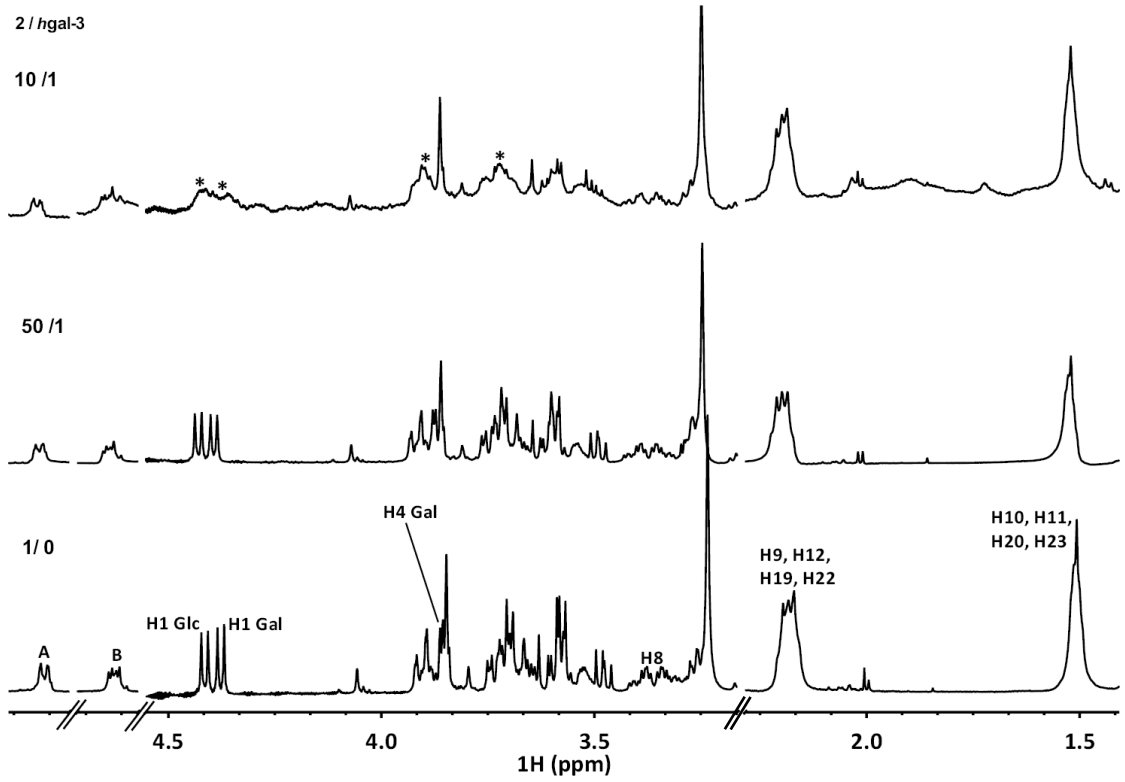


Figure 16. The interaction of 3 with *h*gal-3 CRD. Bottom panel: ^{1}H NMR spectrum for the free ligand in solution. Middle panel: ^{1}H NMR spectrum for a 50:1 molar ratio (2.5 mM:50 μM) mixture of 3:*h*gal-3 CRD. Top panel: ^{1}H NMR spectrum for a 10:1 molar ratio (0.5 mM:50 μM) mixture of 3:*h*gal-3 CRD. The asterisks indicate the ligand protons that show major line broadening. They especially correspond to the Gal moiety. All NMR spectra were processed with the same line broadening functions.

The intermolecular interaction was further confirmed by STD-NMR experiments. The STD (figure 17) spectrum of compound **3** in the presence of *h*gal-3 CRD (1:50 protein:ligand molar ratio) revealed that the terminal lactose moiety displayed the strongest STD effect. The aliphatic protons of the linkers resonating al higher field (H26, H26, H10, H11) also showed STD, while other, some protons far from the sugar moiety (C, H16, H17, H18 and H15) also showed some saturation effects. A control STD-NMR experiment performed or the ligand in absence of the lectin performed under the same irradiation conditions (on resonance frequency at -0.22 ppm) also showed some signals (at δ 3.25, 2.20, and 1.52 ppm), indicating the existence of ligand-ligand interactions. However, the H15-H18 signals did not appear in this spectrum, suggesting the existence of transient non-specific interactions between this ligand region and the lectin.

The STD results reflect that lactose is the major epitope, especially involving H4, H5 and H6 of the terminal Gal moiety, as frequently observed for the interaction of Gal units with galectins.[35]

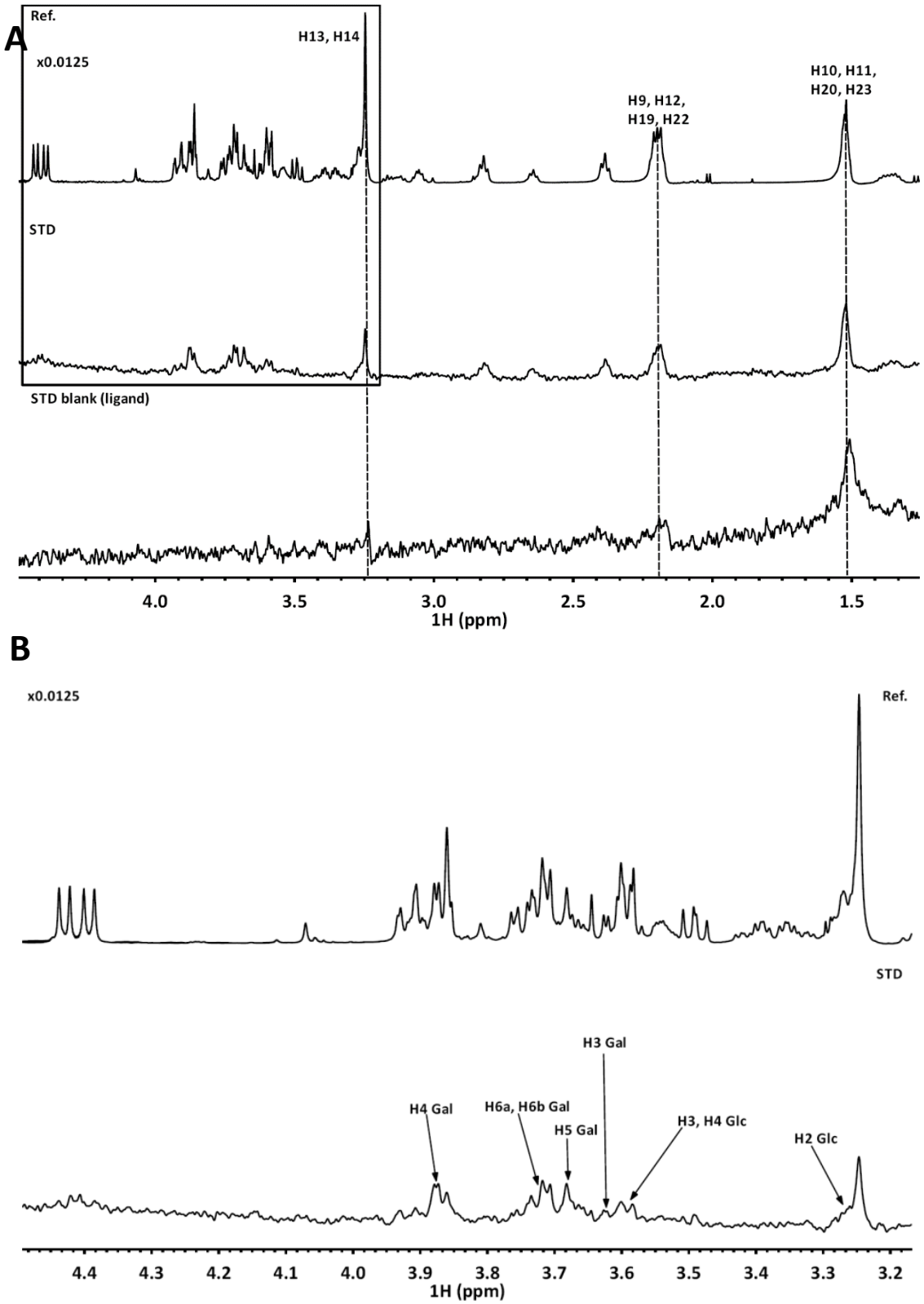


Figure 17. The interaction of 3 with *h*gal-3 CRD. STD-NMR spectra (500 MHz, 298 K) obtained for a sample containing of *h*gal-3 CRD and 3 in a 1:50 molar ratio (50 μM:2.5 mM). A: The spectrum at the bottom show the blank STD spectrum obtained for the free ligand in solution. The middle panel shows the off-resonance spectrum. The top panel shows the STD spectrum. The dashed lines indicate the ligand protons that show STD in the blank experiment. B: Expansion of the spectral region from δ 3.1-4.5 ppm, where the key sugar protons are annotated.

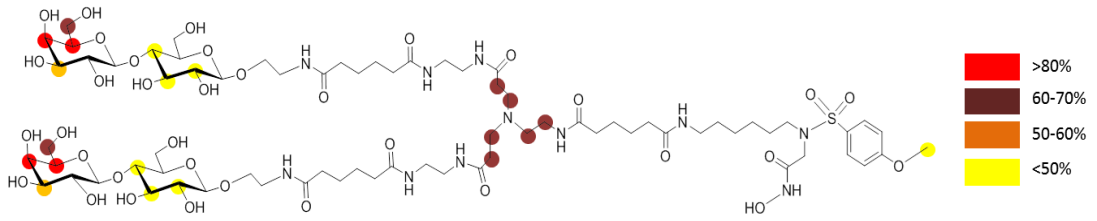


Figure 18. The ligand epitope of compound 3 interacting with *h*gal-3 CRD according to the STD-NMR data.

In order to further asses the carbohydrate-dependent specificity of the interaction between **3** and *h*gal-3 CRD, STD-NMR based competition experiments were pursued. Since the simplest competitor ligand, lactose, could not be used due to the unavoidable signal overlapping with the resonances of the two lactose moieties present in compound **3**, a ^{19}F NMR-based competition study was employed using a fluorinated lactose derivative, namely N-trifluoroacetyl lactosamine, provided by Dr. Stefan Oscarson, at the University of Dublin. This molecule avoids the ^{1}H NMR signal overlapping by taking advantage of the single ^{19}F signal of the competitor. In particular, ^{19}F transverse relaxation rates ($1/T_2$) are very sensitive to changes between the free and bound forms,[65,66] and may be employed to study the competition process. First, the T_2 of the ^{19}F NMR signal of free N-trifluoroacetyl lactosamine was measured (0.5 seconds, figure 19) through the Carr-Purcell-Meiboom-Gill (CPMG) sequence experiment.[67] In the presence of *h*gal-3 CRD (at a ligand:lectin molar ratio of 20:1), the T_2 for the ^{19}F NMR signal was significantly shorter (figures 19 and 20), the linewidth of the ^{19}F signal was significantly larger, and the signal intensity was concomitantly reduced, as a result of the interaction of the disaccharide with the protein. Then, to this solution (N-trifluoroacetyl lactosamine:*h*gal-3 CRD 20:1), compound **3** was added in a stepwise manner and the recovery of the intensity of the ^{19}F NMR signal was recorded each step. It is evident that, the intensity of the ^{19}F NMR signal (figure 20) was systematically recovered approaching the intensity observed for the free disaccharide, indicating that compound **3** is indeed able to remove N-trifluoroacetyl lactosamine from the lectin´s binding site.

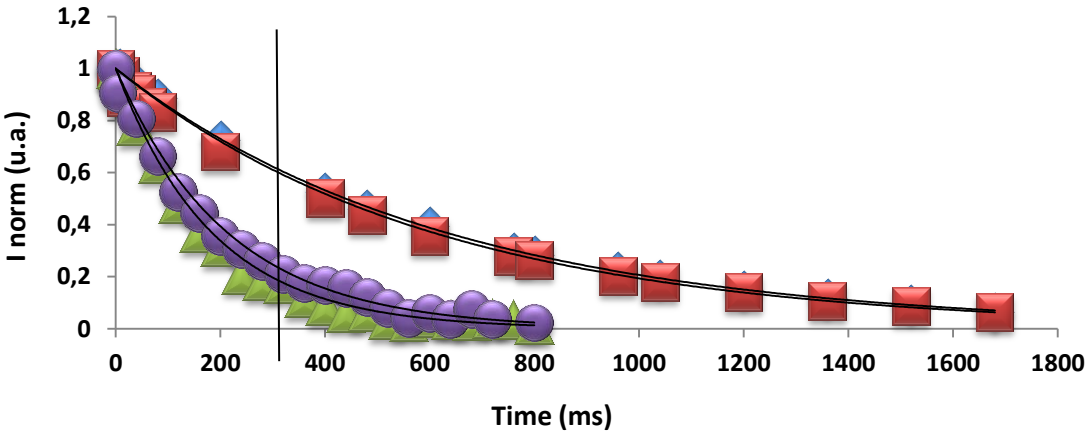


Figure 19. The interaction of N-trifluoroacetyl lactosamine with *h*gal-3 CRD as deduced from CPMG spin echo-based NMR measurements. In red (squares): Intensity decay with the echo time of the ^{19}F NMR signal for the β anomer. In blue (diamond): Intensity decay with the echo time of the ^{19}F NMR signal for the α anomer. In purple (circles): Intensity decay with the echo time of the ^{19}F NMR signal for the β anomer in the presence of *h*gal-3 CRD. In green (triangles): Intensity decay with the echo time of the ^{19}F NMR signal for the α anomer in the presence of *h*gal-3 CRD.

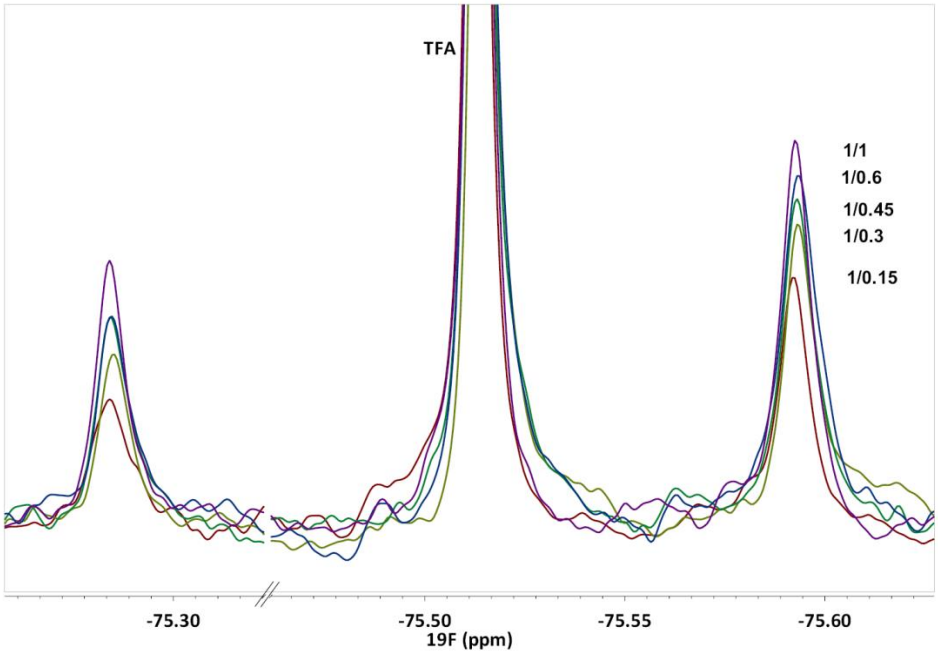


Figure 20 Compound 3 competes with N-trifluoroacetyl lactosamine for the binding site of *h*gal-3 CRD. Representation of the recovery of the ^{19}F NMR signal intensity of N-trifluoroacetyl lactosamine in the CPMG NMR spectra recorded for a solution of N-trifluoroacetyl lactosamine (1 mM) in the presence of 50 μM of *h*gal-3 CRD. The variations of the ^{19}F NMR intensities for the signals of both anomers upon increasing additions of compound 3 are evident (0.5 mM TFA was added as the ^{19}F NMR reference signal).

As further step, the stoichiometry of the binding process was assessed since the dilactosyl branching of the ligand could support the simultaneous binding of two *h*gal-3 CRD. Thus, diffusion ordered spectroscopy (DOSY) was the employed NMR experimental approach to estimate the possible changes in the average hydrodynamic translational diffusion coefficients

in the free and bound states.[68] The diffusion coefficient of *h*gal-3 CRD was monitored by DOSY measurements in the presence of increasing amounts of compound **3**. At sub-stoichiometry amounts of the ligand, where the stoichiometry 2:1 protein:ligand could be favored, no significant changes in the diffusion coefficient of the protein were observed (figure 21). Thus, no evidences of simultaneous binding of two galectin moieties to the same ligand molecule could be obtained, strongly suggesting that under the experimental conditions used, no positive cooperativity (if any occurs) for binding of a second protein was appreciated.

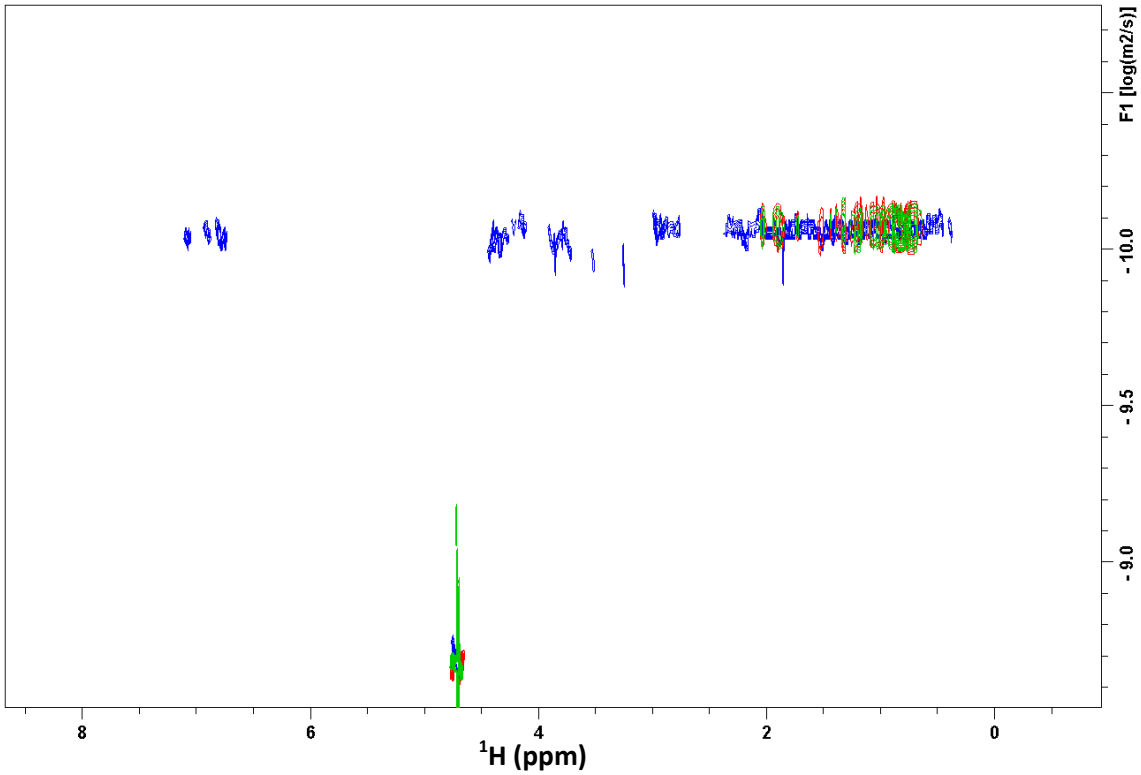


Figure 21. DOSY experiments for *h*gal-3 CRD in the absence and presence of 3. In green, the free *h*gal-3 CRD (200 μM); in red, the mixture of *h*gal-3 CRD (200 μM) and 3 (60 μM); in blue, the mixture of *h*gal-3 CRD (200 μM) and 3 (100 μM). No significant variations in the diffusion coefficient may be appreciated.

Next, a chemical shift perturbation analysis (CSP) based on ^{1}H-^{15}N HSQC experiments was carried out in order to assess that the ligand binds at the canonical lactose-binding pocket of *h*gal-3 CRD. Thus, ^{1}H-^{15}N HSQC spectra (figure 22) were recorded for the ^{15}N labelled *h*gal-3 CRD in the absence and presence of compound **3** (protein:ligand molar ratio 1:10; 0.1 mM:1 mM).

The assignment of the chemical shifts of the backbone NH cross peaks followed the previously described assignment for the free protein (BMRB[69] entry 19491). Chemical shift perturbations after the addition of the bifunctional compound **3** were observed and monitored, showing that major CSP took place around the lactose binding site as defined in the X-ray structures (figure 22) deposited in the Protein Data Bank (PDB[70] ID: 3ZSJ).

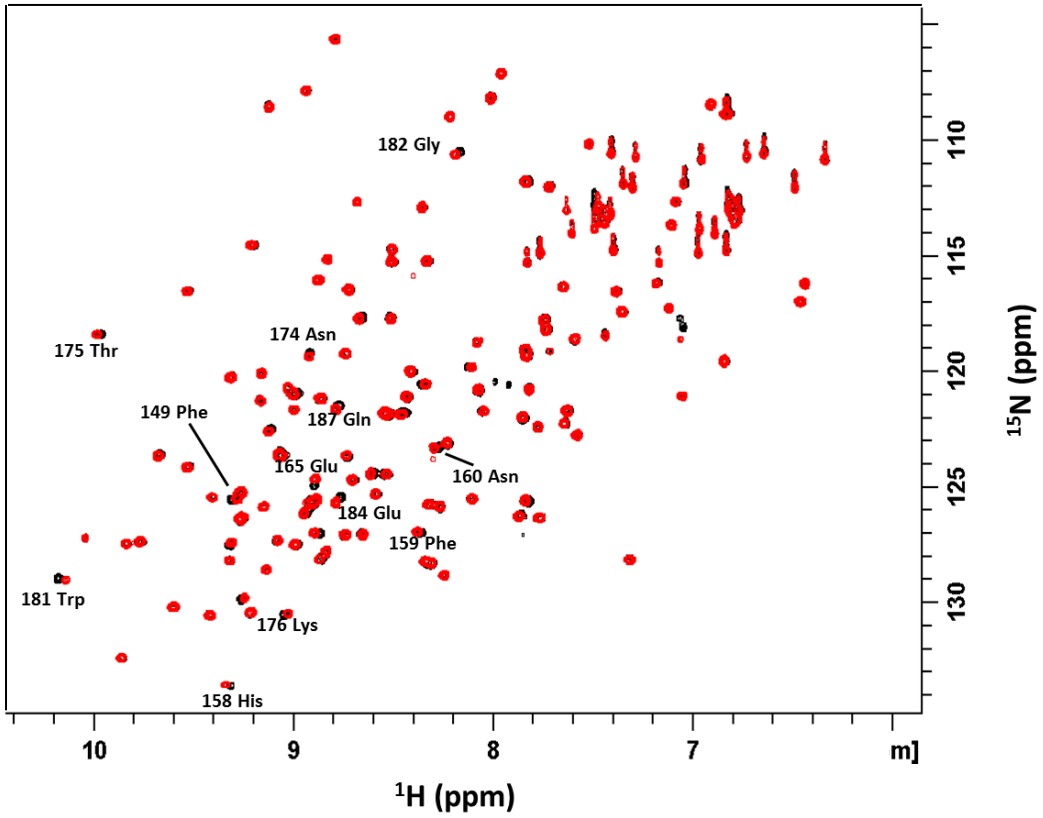


Figure 22. The interaction of 3 with *h*gal-3 CRD. Superimposition of the ^{1}H-^{15}N HSQC spectra recorded for *h*gal-3 CRD in the absence (red) and presence of 1 mM of compound 3 (black).

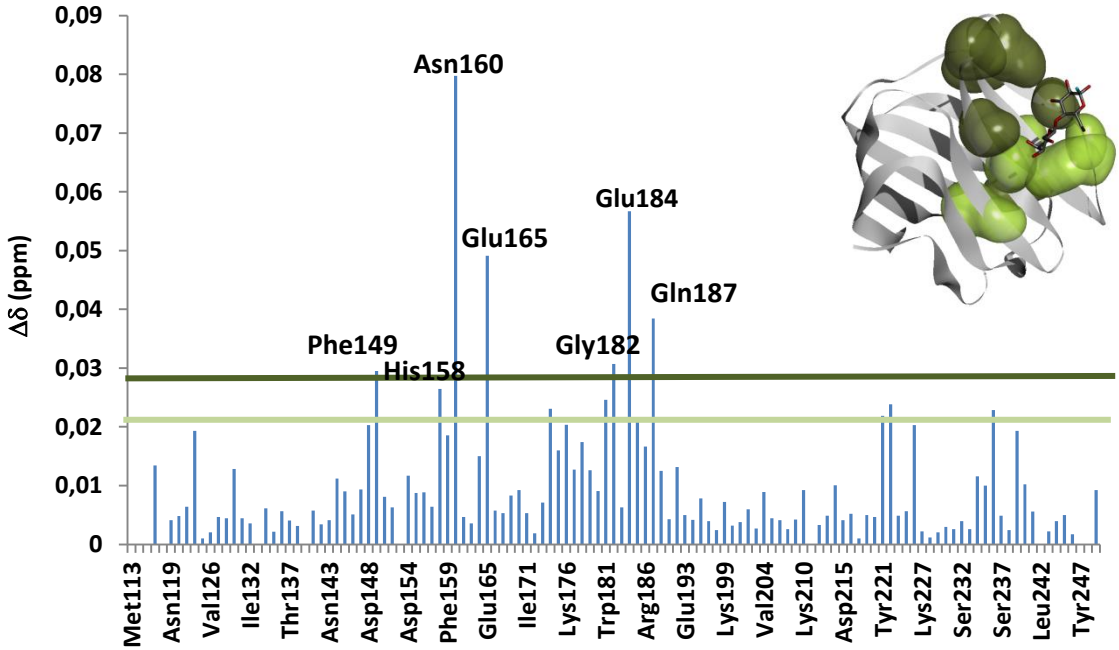


Figure 23. Weighted average CSP of the NH backbone resonances (see experimental section) of *h*gal-3 CRD in the presence of 3. The X-ray structure (PDB ID: 3ZSJ) of *h*gal-3 CRD is shown at the right-hand side. The amino acids with significant CSP of their NH backbone signals are highlighted in surface representation (dark green for those with Δδ above 0.03 ppm and light green for those above 0.02. They correspond to the canonical lactose binding site.

3.3.2. The interaction with the Matrix Metalloproteinase (MMP-12)

Once the interaction towards the *h*gal-3 CRD was assessed, the binding process towards catMMP-12 moiety was also explored. The same methodological NMR protein-based approach was followed, using ^{1}H-^{15}N HSQC-based CSP analysis with a ^{15}N isotopically labelled catMMP-12[71] (BMRB[69] 6391) to investigate the formation of the ligand/protein complex.

The ^{1}H-^{15}N HSQC spectrum of ^{15}N labelled catMMP-12 (159 amino acids) was assigned and compared to that recorded after addition of an excess of compound **3** (figure 24). It should be mentioned that the initial sample of catMMP-12 contained acetohydroxamic acid (AHA) in order to inhibit the intrinsic proteolytic activity and to ensure that the protein life time is enough to record good quality ^{1}H-^{15}N HSQC spectra.[71]

Figure 25 shows the weighted average CSP of the catMMP-12 NH backbone resonances upon addition of **3.** Three different regions of catMMP-12 that are clearly perturbed when the inhibitor AHA is displaced by **3**, a stronger binder. These regions can be defined as the initial part of beta strand IV (G178-H183), T215 in the alpha 2 helix closed to the catalytic site, and the so called S1´pocket (S229-Y242), where the aromatic moiety is supposed to fit.

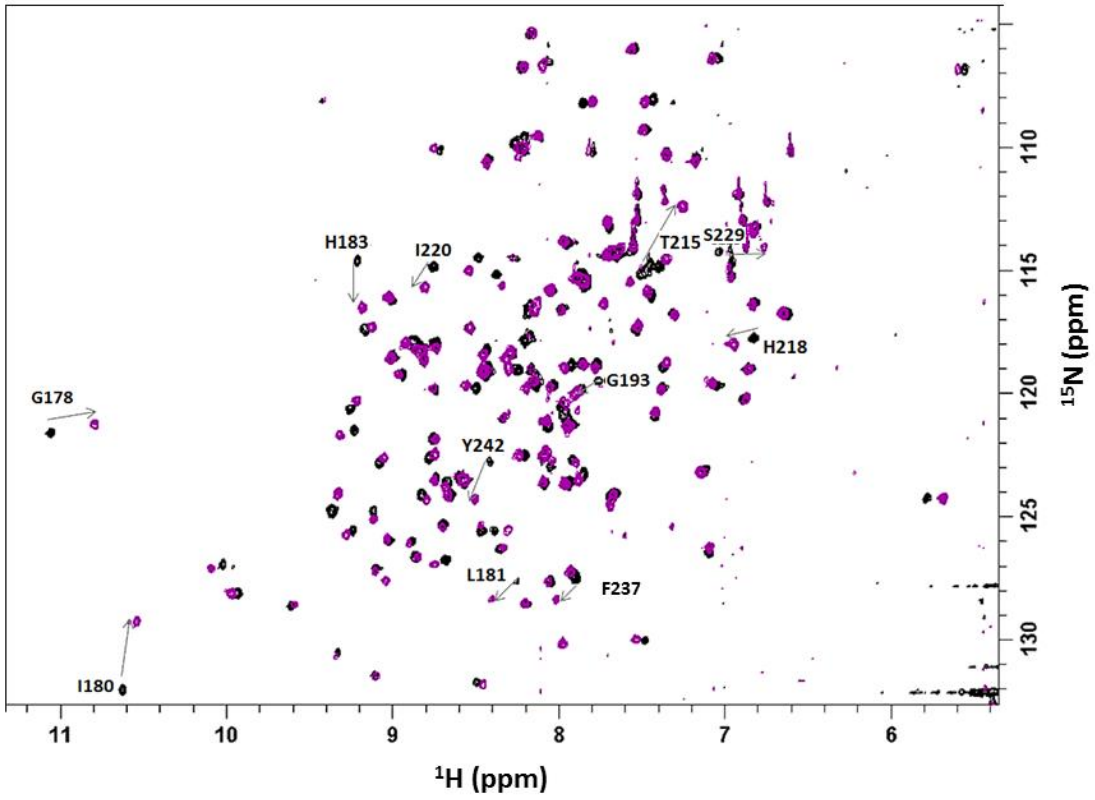


Figure 24. The interaction of 3 with catMMP-12. Superimposition of the ^{1}H-^{15}N HSQC spectra recorded for ^{15}N-catMMP-12 in the presence of AHA (black) and after the addition of compound 3 (purple). The key perturbed cross peaks are highlighted

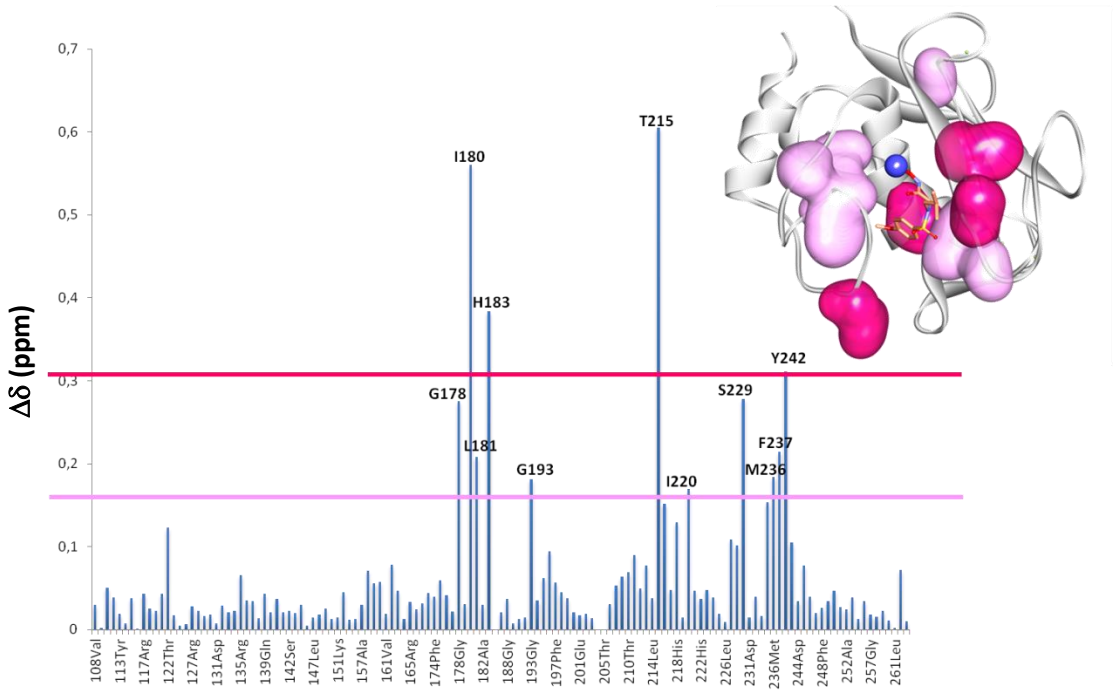


Figure 25. Weighted average CSP of the catMMP-12 NH backbone resonances upon interaction with 3 (with respect to the chemical shifts in the presence of AHA). In the right-hand side, the X-ray structure (PDB ID:1Z3J) of the catMMP-12 is shown. The amino acids with the strongest chemical shift perturbations in their backbone NH are highlighted in the surface representation in red for those with Δδ above 0.3 ppm and in pink for those over 0.17 ppm).

These observations are in full accordance with the binding mode previously described[64] for MMP inhibitors based on N-isobutyl-N-(4-methoxyphenylsulfonyl) glycyl hydroxamic acid (NNGH). Therefore, the presence of the lactose moiety does not interfere with the binding of the hydroxamic moiety to catMMP-12.

3.3.3. The *h*gal-3 CRD:**3**:catMMP-12 ternary complex

After these individual studies of the binding of compound **3** to each single protein, the formation of the *h*gal-3 CRD:**3**:**c**atMMP-12 ternary complex was evaluated. In this case, ^{1}H-^{15}N HSQC experiments focused on the observation of the catMMP-12 signals were performed.

First, the ^{1}H-^{15}N HSQC spectrum of catMMP-12 in the presence of an excess of compound **3** was recorded as a control. Under these conditions, the catMMP-12 binding site is saturated with the ligand. Second, an equimolecular amount of *h*gal-3 CRD with respect to catMMP-12 was added to the solution and the corresponding HSQC spectrum was recorded under the same experimental conditions. The comparison of the ^{1}H-^{15}N HSQC spectra recorded before and after the addition of *h*gal-3 CRD showed a clear decrease in intensities of many NH cross peaks (figure 26). This observation is likely due to the change in apparent molecular weight of

the observed catMMP-12 due to the formation of supramolecular complex, with the concomitant increase of the transverse relaxation rate and reduction of the intensity of the ^{1}H-^{15}N HSQC cross-peaks. We can therefore assume that a ternary complex is formed, in which **3** is able to simultaneously bind to catMMP-12 through the hydroxamic acid functionality and to *h*gal-3 CRD through the lactose epitope.

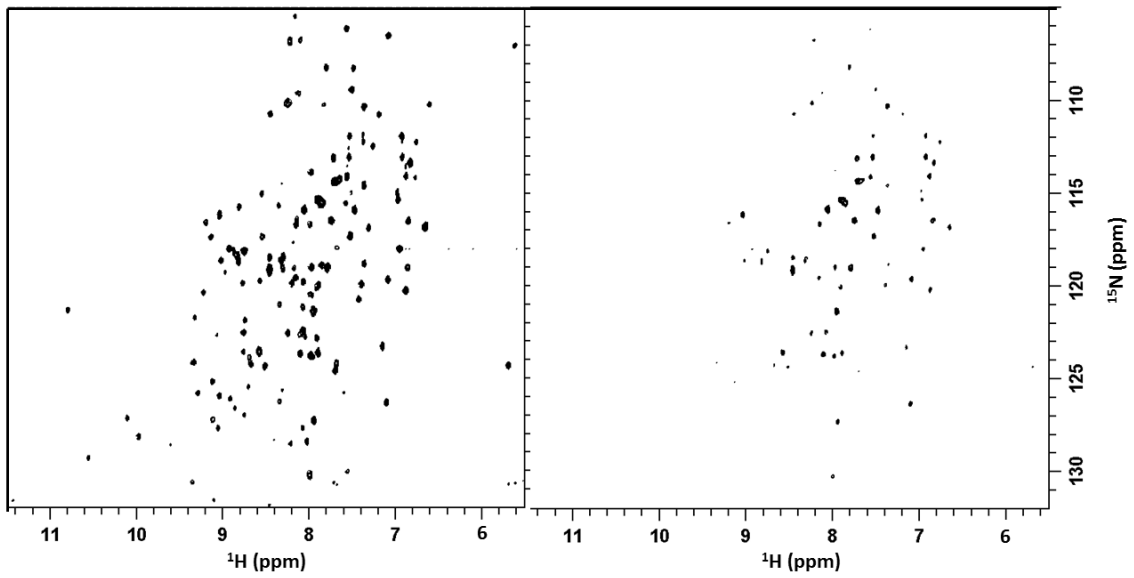


Figure 26. The formation of the ternary complex through the measurement of ^{1}H-^{15}N HSQC spectra for catMMP-12. On the right-hand side, ^{1}H-^{15}N HSQC spectrum recorded for the mixture of catMMP-12 and 3 (in molar excess), without the lectin. On the left-hand side, ^{1}H-^{15}N HSQC spectrum recorded for the same mixture after the addition of an equimolar amount of *h*gal-3 CRD (with respect to catMMP-12).

3.3.4. Conclusions

A novel bifunctional ligand to simultaneously interaction with two different protein targets has been synthesized. The ligand shows a divalent display of lactose epitopes to bind galectins and a hydroxamic moiety to target matrix metalloproteinases. The NMR investigation has permitted to assess, in a non-ambiguous manner, the existence of interaction between the designed epitopes with their corresponding proteins in a similar manner to that taking place for the individual portions, while evidences for the formation of a ternary complex between the ligand and the two proteins has also been derived from the NMR analysis.

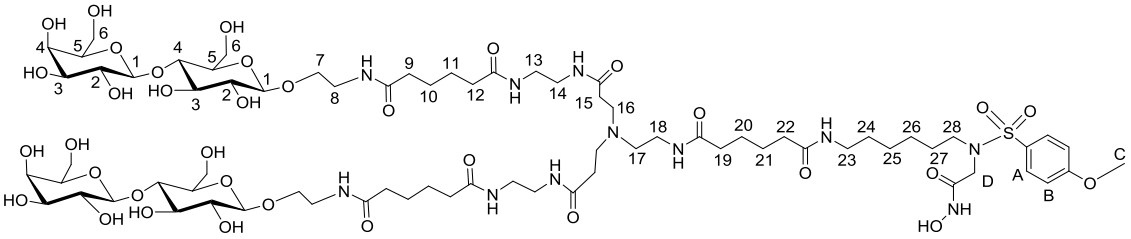
3.3.5. Experimental

The bifunctional compound **3** was synthesized by the group of Prof. Cristina Nativi at the University of Florence. The synthesis will be published elsewhere. The N-trifluoroacetyl lactosamine was provided by Prof. Stefan Oscarson at the University of Dublin.

The *h*gal-3 CRD samples (^{15}N isotopically labelled and unlabeled carbohydrate recognition domain from amino acids Met113 to Met249) were provided by Dr. Filipa Marcelo from the University Nova of Lisbon. The assignment was taken from BMRB entry 14941.

The catMMP-12 ^{15}N isotopically labeled (V108 to G263) was provided by Marco Fragai from University of Florence.[71] The assignment was taken from BMRB 6391.

Ligand resonances were assigned by the acquisition of a set of standard NMR experiments: 1D ^{1}H-NMR, 2D homonuclear spectra TOCSY at different mixing times and ROESY and 2D heteronuclear ^{1}H-^{13}C HSQC in D_2O in a 600 MHz Bruker Avance spectrometer equipped with a cryoprobe.



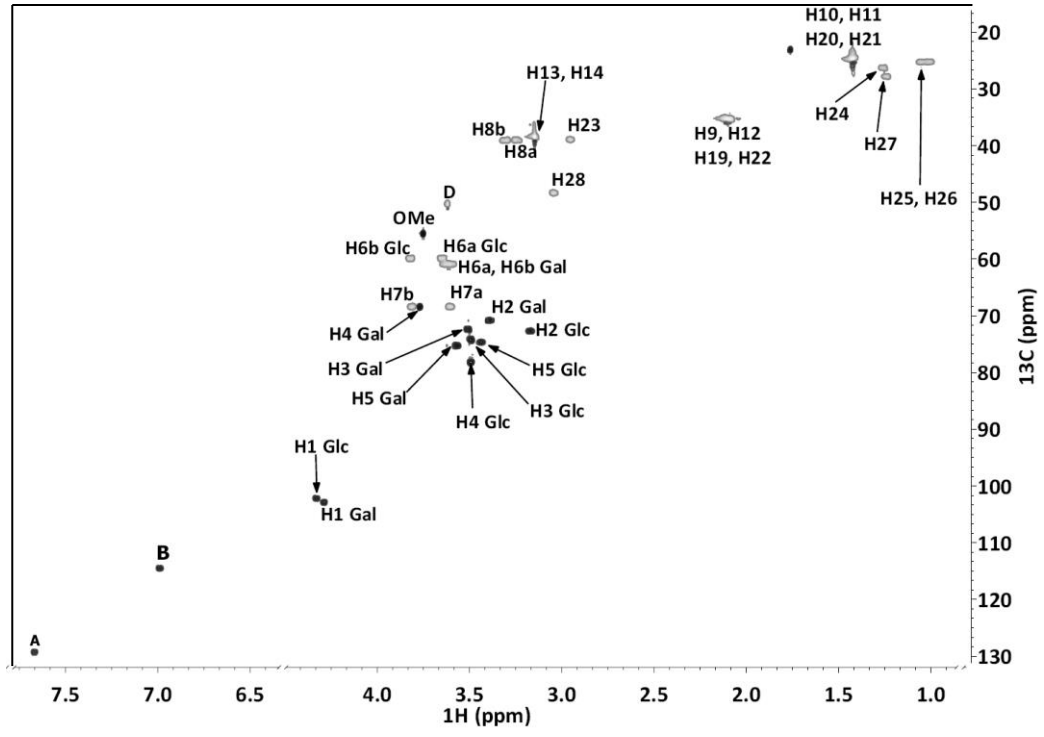


Figure 27. ^{1}H-^{13}C HSQC of compound 3.

The chemical shift data of 3 are listed in the following table.

Proton	^{1}H-Chemical Shift (δ ppm)	^{13}C-Chemical Shift (δ ppm)
H1 Gal	4.29	102.9
H2 Gal	3.39	70.8
H3 Gal	3.50	72.4
H4 Gal	3.77	68.4
H5 Gal	3.57	75.3
H6a Gal	3.60	60.9
H6b Gal	3.60	60.9
H1 Glc	4.33	102.1
H2 Glc	3.17	72.7
H3 Glc	3.48	74.2
H4 Glc	3.49	78.2
H5 Glc	3.43	74.7
H6a Glc	3.64	59.9
H6b Glc	3. 81	59.9
H7a	3.60	68.4
H7b	3.80	68.4
H8a	3.25	39.1
H8b	3.30	39.1
H9	2.09	35.3
H10	1.42	24.7
H11	1.42	24.7
H12	2.09	35.3
H13	3.15	38.4
H14	3.15	38.4
H15	2.28	-
H16	2.68	-
H17	2.49	-
H18	2.28	-
H19	2.11	35.3
H20	1.42	24.7
H21	1.42	24.7
H22	2.09	35.3
H23	2.95	38.9
H24	1.25	26.3
H25	1.05	25.3
H26	1.00	25.2
H27	1.24	27.9
H28	3.04	48.4
A	7.66	129.3
B	6.98	114.5

Table 5. Chemical shift of the ^{1}H resonances of 3 at 298 K (600 MHz).

For the line broadening analysis, ^{1}H-NMR spectra were acquired in a 500 MHz Bruker Avance spectrometer equipped with a cryoprobe on a sample containing 50 μM of *h*gal-3 CRD in PBS deuterated buffer pH 7.3. A 0.5 mM ligand concentration was used for the sample at a 1:10 *h*gal-3 CRD:**3** molar ratio with 50 μM of *h*gal-3 CRD, while a 2.5 mM ligand concentration for the sample at the 1:50 *h*gal-3 CRD:**3** ratio. All spectra were processed with the same line broadening function.

STD-NMR experiments. The sample with the highest ratio protein:ligand molar ratio (1:50) was used for the STD-NMR experiments. The on-resonance frequency was set at -0.22 ppm with a total saturation time of 2 seconds and a spin lock filter to minimize the background of the protein resonances. No water suppression was applied.

^{19}F competition study. A 500 MHz (^{1}H frequency) Bruker Avance spectrometer was employed to record a set of ^{19}F-NMR spectra on a sample containing N-trifluoroacetyl lactosamine (1 mM) in the presence of *h*gal-3 CRD (50 μM) in PBS deuterated buffer (pH 7.2) and TFA (0.5 mM, used as internal standard). Increasing amounts of **3** were added to the sample in stepwise manner and ^{19}F-NMR spectra were recorded after each addition. The final point was recorded for a sample displaying an equimolecular concentration of N-trifluoroacetyl lactosamine and **3** (table 6).

	*h*gal-3 CRD (μM)	N-trifluoroacetyl lactosamine (mM)	Compound 3 (mM)
Exp. 1	50	1	0.15
Exp. 2	50	1	0.30
Exp. 3	50	1	0.45
Exp. 4	50	1	0.60
Exp. 5	50	1	1

Table 6. Experimental conditions of the ^{19}F-NMR competition study between 3 and N-trifluoroacetyl lactosamine.

Diffusion NMR experiments. DOSY experiments were performed using a 500 MHz Bruker Avance spectrometer. The DOSY experiments were acquired with 16 gradient increments to a final intensity decay of 90%. The measured samples contained 200 μM *h*gal-3 CRD in PBS deuterated buffer pH 7.3 in the absence and presence of different amounts of **3** (60 μM and 100 μM).

^{1}H-^{15}N heteronuclear correlation spectra. ^{1}H-^{15}N HSQC spectra were recorded using a 600 MHz Bruker Avance spectrometer equipped with a cryoprobe. For the catMMP-12, spectral widths of δ 14 ppm (centered at δ 4.70 ppm) in the ^{1}H dimension and δ 40 ppm (centered at δ 118

ppm) in the ^{15}N dimension were used. The same parameters were employed for the multimeric complex analysis.

For *h*gal-3 CRD, the spectral width was δ 14 ppm (centered in δ 4.70 ppm) in the ^{1}H dimension and δ 34 ppm (centered in δ 119 ppm at the ^{15}N dimension). For ^{15}N-labeled *h*gal-3 CRD (0.1 mM) two spectra were acquired, either in the absence or in the presence of compound **3** (1 mM), in PBS buffer at pH 7.3. For ^{15}N-labeled catMMP-12 (0.2 mM), two spectra were acquired, either in the absence or in the presence of compound **3** (2 mM) in perdeuterated-TRIS buffer pH 7.9.

The chemical shift perturbations were analyzed by calculating the corresponding weighted average data according to the following equation:

$$\Delta\delta(ppm) = \sqrt{\left(\Delta\delta\left(\,^{1}H\right)\right)^{2} + \left(\frac{\Delta\delta\left(^{15}N\right)}{5}\right)^{2}}$$

For analyzing the formation of the ternary complex, a sample of ^{15}N-labeled catMMP-12 (0.08 mM) in the presence of a slight excess of compound **3** (0.1 mM, corresponding to a 1:1.25 molar ratio) was employed. ^{1}H-^{15}N HSQC of the complex was acquired in the absence and presence of non-labeled *h*gal-3 CRD (0.18 mM) in perdeuterated TRIS buffer at pH 7.9.

3.4. Glycan-antibody interactions: One example

3.4.1. Anti-Gal antibodies

The α-galactose (α-gal) epitope is a unique carbohydrate structure that is absent in humans but is present in numerous animal species.[72] Moreover, all humans produce anti-gal antibodies, as the IgG displayed in all humans or the anti-gal IgM and IgA isotypes.[73] Since the α-gal epitope is an antigen for humans, it triggers the antibody-antigen reaction, thus activating the immune response and the complement system.[74,75] However, the shortage of human organs for transplantation has opened the door to new alternatives involving the use of animal organs for humans; xenotransplantation. In this context, porcine organs have been considered as suitable xenografts although different obstacles have to be avoided. Rejection of xenografts is related to the fast anti-gal antibody/α-gal epitope immune response.[76] The rejection is produced by hyperacute response (HAR), a mechanism that is activated within a few minutes.[74,75,77,78] Moreover, acute humoral xenograft rejection (AXHR) may also appear within hours once HAR is avoided.[78,79] The existence of these two challenging events, HAR and

AXHR, have led to the development of different strategies to bypass the rejection, such as the immunosuppression of the human immunological system or the use of α-gal epitope knock out pig xenografts.[74,77,80] Unfortunately, the clinical efficacy of these strategies, even combined, has not been successful. Indeed, the high sensitivity and avidity of the anti-gal human antibody towards the xenoantigen finally triggers the rejection of the xenograft.[77] Nevertheless, the possibilities of exploiting the anti-gal antibody/α-gal epitope immune response has recently gained increased attention for the development of autologous vaccines[75,81,82] following cancer therapies and then kill the residual cancer cells and avoid metastasis. In the present work, the recognition features of different monoclonal anti-gal antibody fragments have been studied.

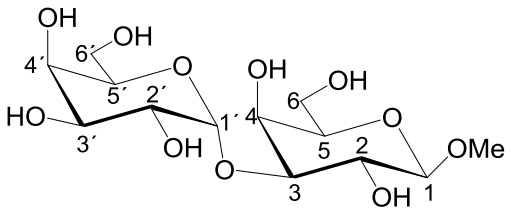


Figure 28. The α-gal epitope disaccharide (Galα1-3Gal)

In particular, three monoclonal antibodies, anti-gal A4, A11 and G12, have been employed. The antibody constructs consist in fragments of their hyper-variable light chains, including the binding sites.

A NMR approach using the ligand point of view was applied to address the recognition of the antibodies and evaluate the ligand epitope and the interaction mode. The commercially available disaccharide Galα-1-3GalβOMe was employed as α-gal epitope probe. After the full assignment of the NMR spectra of the disaccharide using a standard set of 1D and 2D NMR experiments, STD-NMR experiments were carried out, as described in the experimental section.

Figure 29 shows the STD spectrum of the α-gal epitope in the presence of the anti-gal A4 antibody. The analysis showed the interaction of the disaccharide preferentially at the non-reducing end. The protons H3´, H2´, H1´and H5´ received the highest saturation from the antibody. Additionally, the reducing end of the disaccharide also showed STD responses although with lower intensities than those displayed by the non-reducing galactose.

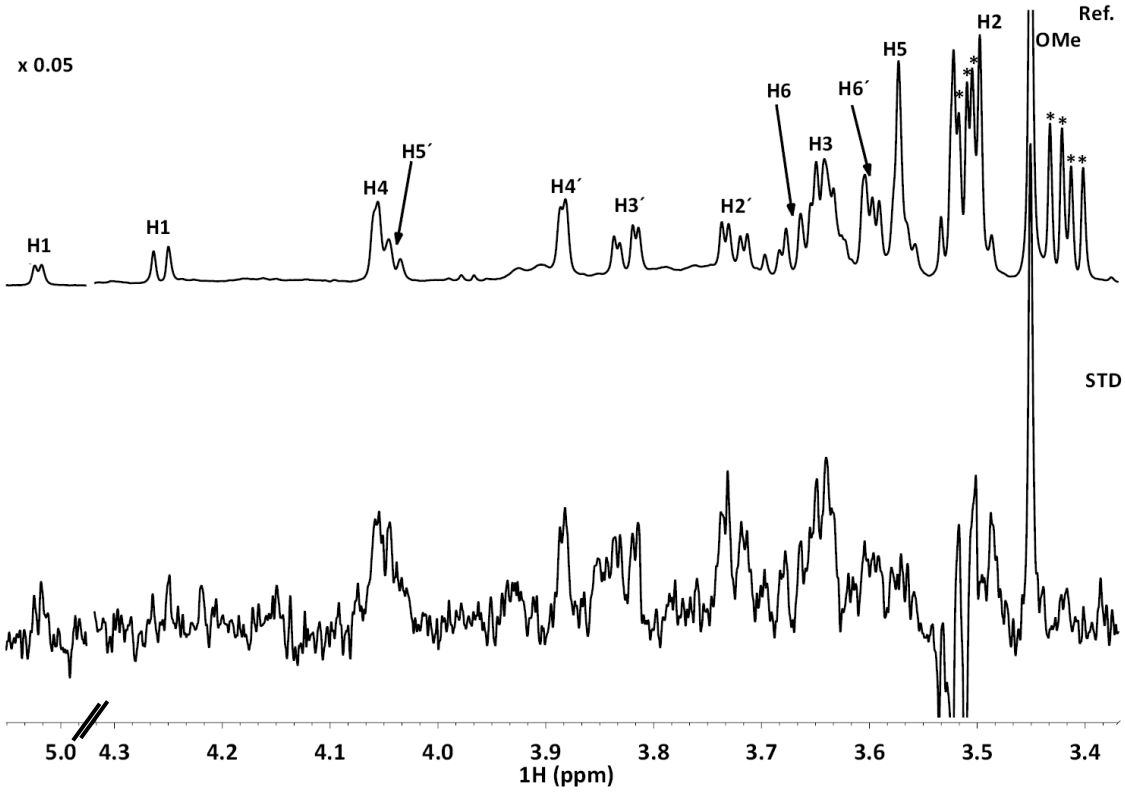


Figure 29. STD (below) and off-resonance (on top) NMR spectra (600 MHz) of the interaction of the disaccharide with the anti-gal A4 antibody. The temperature was 298 K and a 2 seconds saturation time at the aliphatic region (δ 0.8 ppm) was employed. Glycerol impurities present in the sample are marked with stars.

	Absolute STD intensity (%)	Relative STD intensity (%)
H1	3.9	59
H4	3.8	57
H6	3.9	59
H1’	6.7	100
H2'	6.6	99
H3’	5.0	75
H4'	3.8	57
H5'	6.1	92
H6'	2.3	35
OMe	2.5	38

Table 7. The STD intensities of the disaccharide in the presence of the anti-gal A4 antibody.

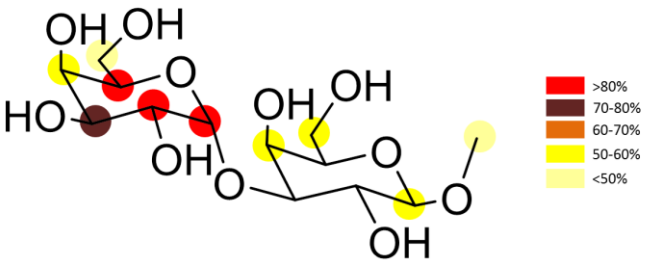


Figure 30. Epitope mapping of the ligand when interacting with the anti-gal A4 antibody.

In contrast, in the presence of the anti-gal A11 antibody, the STD spectrum (figure 31) of the disaccharide showed very low STD intensities, which precluded the accurate quantification of the intensities. H2´ and H3´ received the highest saturation but H1 of the reducing end showed high STD as well. According to these data, the binding epitope includes both saccharide units from the α-gal epitope. Nevertheless, the observed low absolute STD intensities below 1% (table 8) strongly suggest that anti-gal A11 antibody recognizes the α-Gal epitope with significantly lower affinity than the anti-gal A4 analogue.

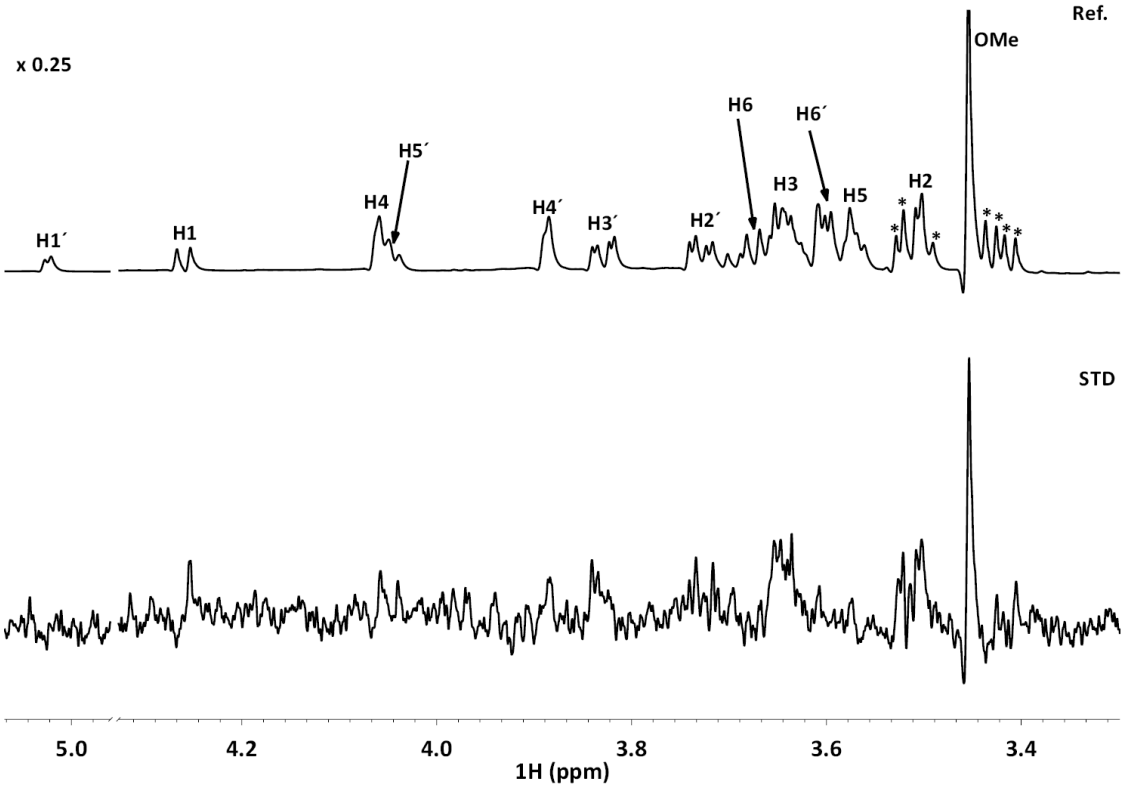


Figure 31. STD (below) and off-resonance (on top) NMR spectra (600 MHz) of the interaction of the disaccharide with the anti-gal A11 antibody. The temperature was 298 K and a 2 seconds saturation time at the aliphatic region (δ 0.8 ppm) was employed. Glycerol impurities present in the sample are marked with stars.

	Absolute STD intensity (%)	Relative STD intensity (%)
H1	0.36	50
H2	0.42	58
H3	0.46	64
H4	0.32	44
H2'	0.72	100
H3'	0.72	100
H4'	0.28	39
H5'	0.37	51
OMe	0.12	17

Table 8. The STD intensities estimated for the disaccharide in the presence of the anti-gal A11 antibody.

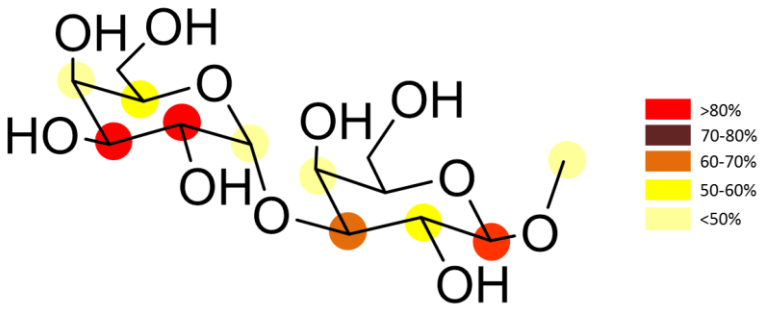


Figure 32. Epitope mapping analysis of the ligand in the presence if the anti-gal A11 antibody.

Finally, similar STD results to those described above for the A11 antibody were obtained for the disaccharide in the presence of the anti-gal G12 analogue (figure 33). In this case, the measured STDs are rather low, precluding the accurate measurement of the STD intensities.

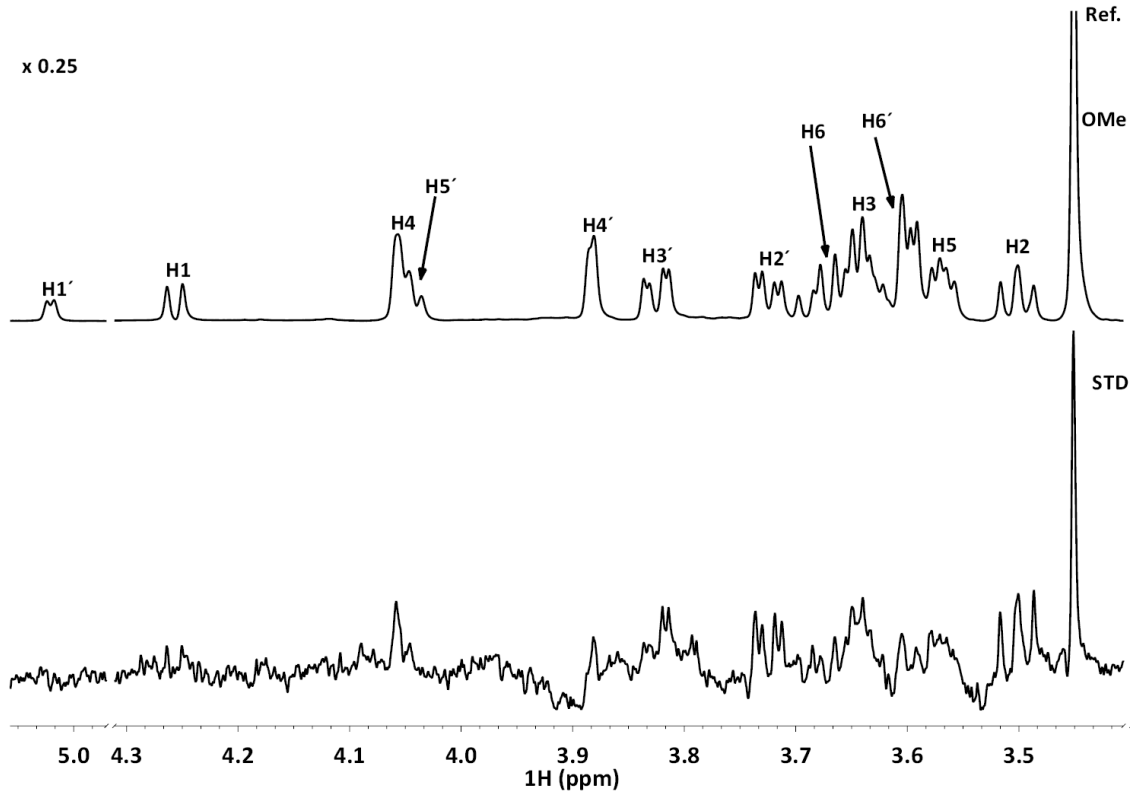


Figure 33. STD (below) and off-resonance (on top) NMR spectra (600 MHz) of the interaction of the disaccharide with the anti-gal G12 antibody. The temperature was 298 K and a 2 seconds saturation time at the aliphatic region (δ 0.8 ppm) was employed.

	Absolute STD intensity (%)	Relative STD intensity (%)
H1	0.7	56
H2	1.1	91
H4	0.6	50
H2'	1.2	100
H3'	0.8	69
H5'	0.5	42
H6'	0.2	18
OMe	0.5	43

Table 9. The STD intensities estimated for the disaccharide in the presence of the anti-gal G12 antibody.

The epitope mapping of the α-gal epitope shows that H2´and H3´received the highest saturation as well as H2 from the reducing end.

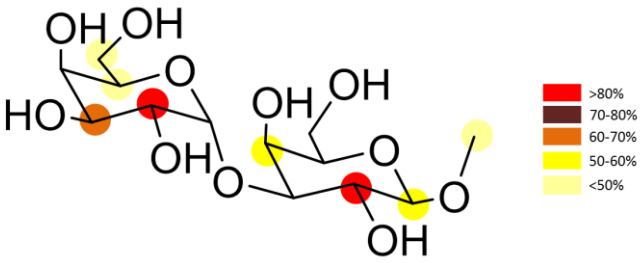


Figure 34. Epitope mapping analysis of the ligand in the presence of the anti-gal G12 antibody.

3.4.2. Conclusions

The analysis of the STD data for the α-gal epitope disaccharide (Galα1-3Gal) in the presence of the three monoclonal antibodies permitted to demonstrate that all of them indeed bind the xenoantigen (α-gal) by recognizing the non-reducing end, especially at the Gal H2-H3 region. However, the STD intensities differ among the three antibodies and suggest that the anti-gal A4 displays the strongest binding for Galα1-3Gal while anti-gal A11 and G12 recognize the α-gal epitope disaccharide with similar affinity, although significantly lower than the anti-gal A4 antibody.

3.4.3. Anti-sialic acid antibodies

It is noteworthy to mention that artificial monoclonal antibodies are not only useful for clinical applications but they may also be employed for epitope analysis using histopathology. In fact, immunohistochemistry takes advantage of antibody-antigen interactions to detect key epitopes present at the cell membrane or at the surface of the tissues.[83–85] As already mentioned, a high diversity of glycoproteins and glycolipids cover the cell membrane forming the glycocalix. Using the glycocalix, the high diversity of glycosylations plays an important role in extracellular cell events such as cell-cell communication,[86] virus infection[85] or immune response.[9,75,87] In humans, N-glycans, O-glycans, and glycosphingolipids are often terminated with a sialic acid, a nine carbon atoms monosaccharide with a carboxylic moiety. Sialic acid also display structural diversity including the N-acetylneuraminic (Neu5Ac) and N-glycolylneuraminic (Neu5Gc) acids among those with biological significance.[88–91] In humans, Neu5Ac is the sialic acid derivative most abundant in the respiratory pathways and is involved in influenza infection.[92–94] In contrast, humans are not able to produce Neu5Gc, but this derivative may originate from the dietary incomes or can be metabolically incorporated.[95,96]

Despite the importance of sialic acids, their exact roles still remain uncertain. However, it is today evident that their dysregulations are associated to medical disorders.[90]

In this context, we have also investigated the interaction of anti-sialic antibodies, anti-sialic **1** and anti-sialic **2**, with the corresponding epitopes. These antibodies have been proposed as probes for immunohistochemical studies to try to detect sialic acid-based disorders. Two different fragments of the hyper-variable light chain of the antibody have been expressed that specifically target sialic acid derivatives. Again, NMR methodology based on the ligand perspective has been adopted to monitor the corresponding interactions.

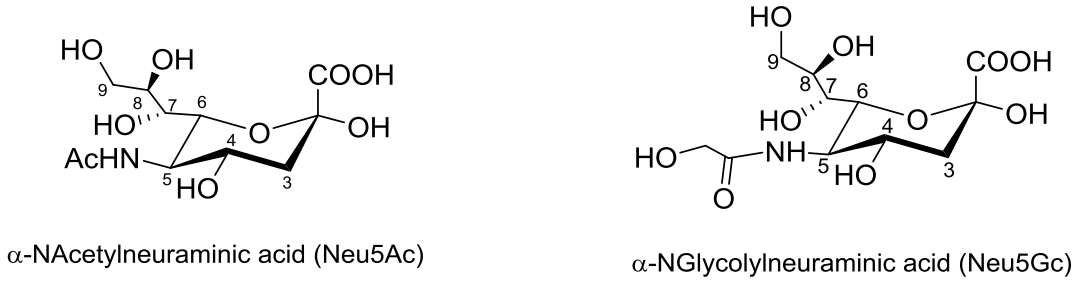


Figure 35. Chemical structures of sialic acid derivatives Neu5Ac (left) and Neu5Gc (right).

The STD-NMR analysis of Neu5Gc in the presence of anti-sialic **1** (figure 37) showed the preferred recognition of the minor anomer present in solution, the α anomer. Indeed, H9, H5 and H7 from the α-anomer displayed the highest STD intensities while the predominant β-anomer showed much smaller saturation (table 10).

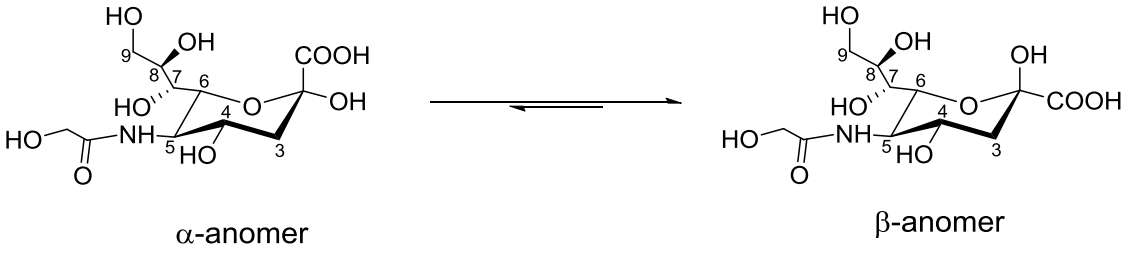


Figure 36. The mutarrotation equilibrium between the Neu5Gc anomers. The intermediate open form is not shown.

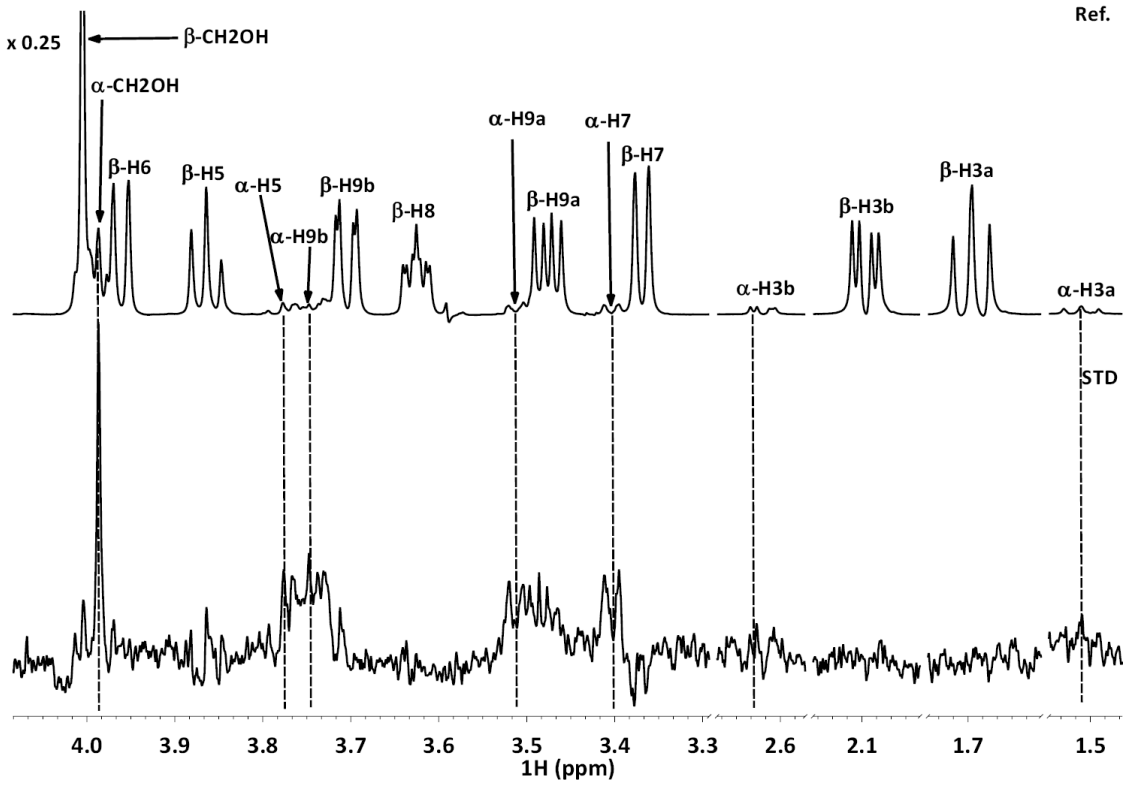


Figure 37. STD (below) and off-resonance (on top) NMR spectra of the interaction of Neu5Gc with the anti-sialic 1 antibody. The temperature was 298 K and a 2 seconds saturation time at the aromatic region (δ 7 ppm) was employed.

	Absolute STD intensity (%)	Relative STD intensity (%)
α-H3a	4.16	60
α-H3b	4.20	60
α-H5	6.64	95
α-H7	6.90	99
α-H9a	6.32	91
α-H9b	6.97	100
α- CH_2OH	3.29	47
β-H5	0.35	5
β-H9a	0.49	7
β-H9b	0.38	5
β-CH_2OH	0.10	1

Table 10. The STD intensities estimated for Neu5Gc in the presence of the anti-sialic 1 antibody.

The STD results show the highly-preferred binding towards the α-anomer. The importance of the anomer configuration for the binding event towards the anti-sialic **1** antibody may have biological relevance. In humans, sialic acid is often found attached to the glycan using α configuration, although the β anomer accommodates the bulky carboxylate group in an equatorial orientation. Thus, the specificity of the anti-sialic **1** antibody could be exploited to detect α linked Neu5Gc moieties in human tissues.

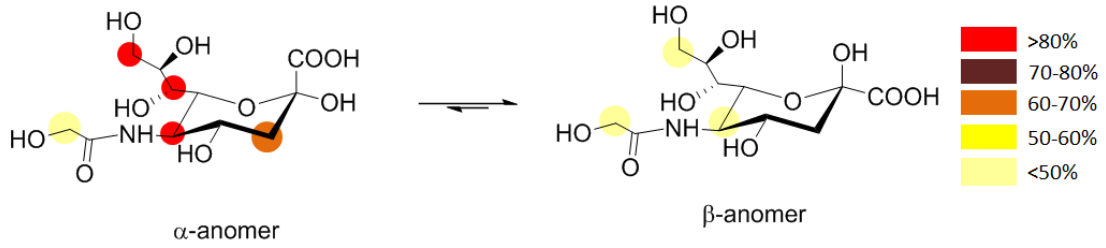


Figure 38. Epitope mapping analysis of Neu5Gc in the presence of the anti-sialic 1 antibody.

Fittingly, no STD was observed for the Neu5Ac analogue in the presence of the anti-sialic 1 antibody under the same experimental conditions employed for the Neu5Gc analogue (figure 39).

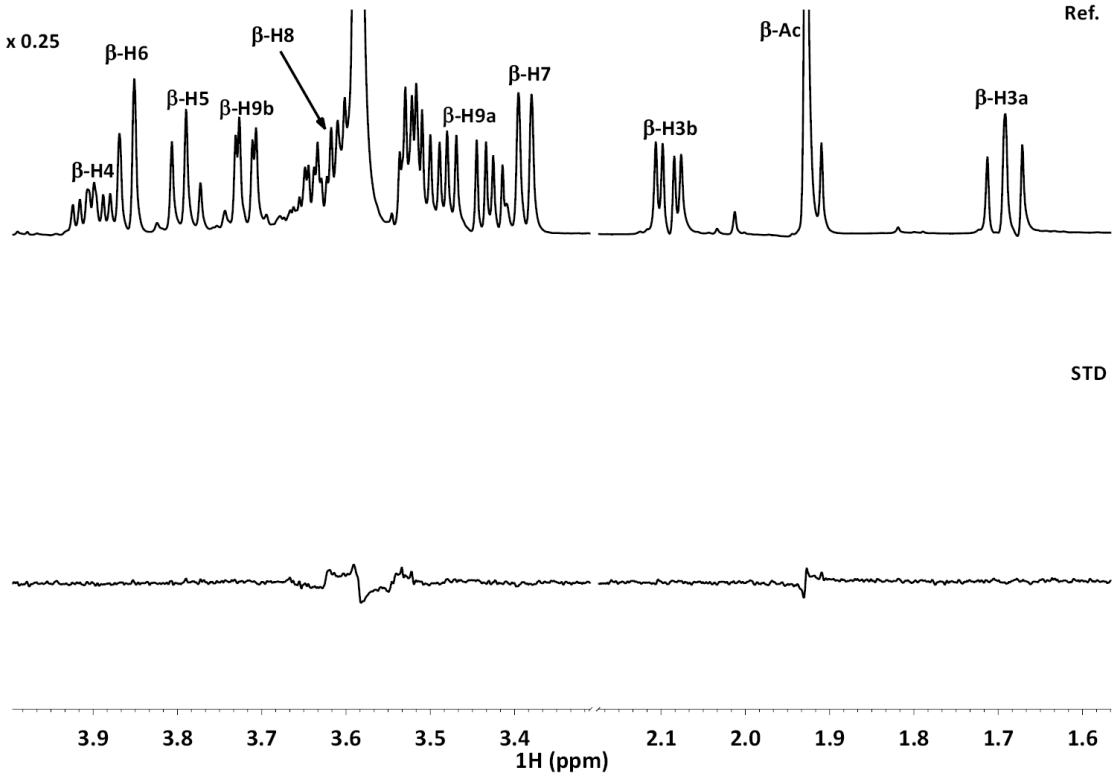


Figure 39. No STD intensities (bottom) are observed for Neu5Ac in the presence of the anti-sialic 1 antibody. The off-resonance spectrum is shown on the top panel.

Next, the interaction of the anti-sialic **2** antibody against Neu5Gc and Neu5Ac was studied. Again, a high specificity towards the α anomer of Neu5Gc was observed (figure 40) with the region defined by H9, H7 and H5 composing the key interacting epitope.

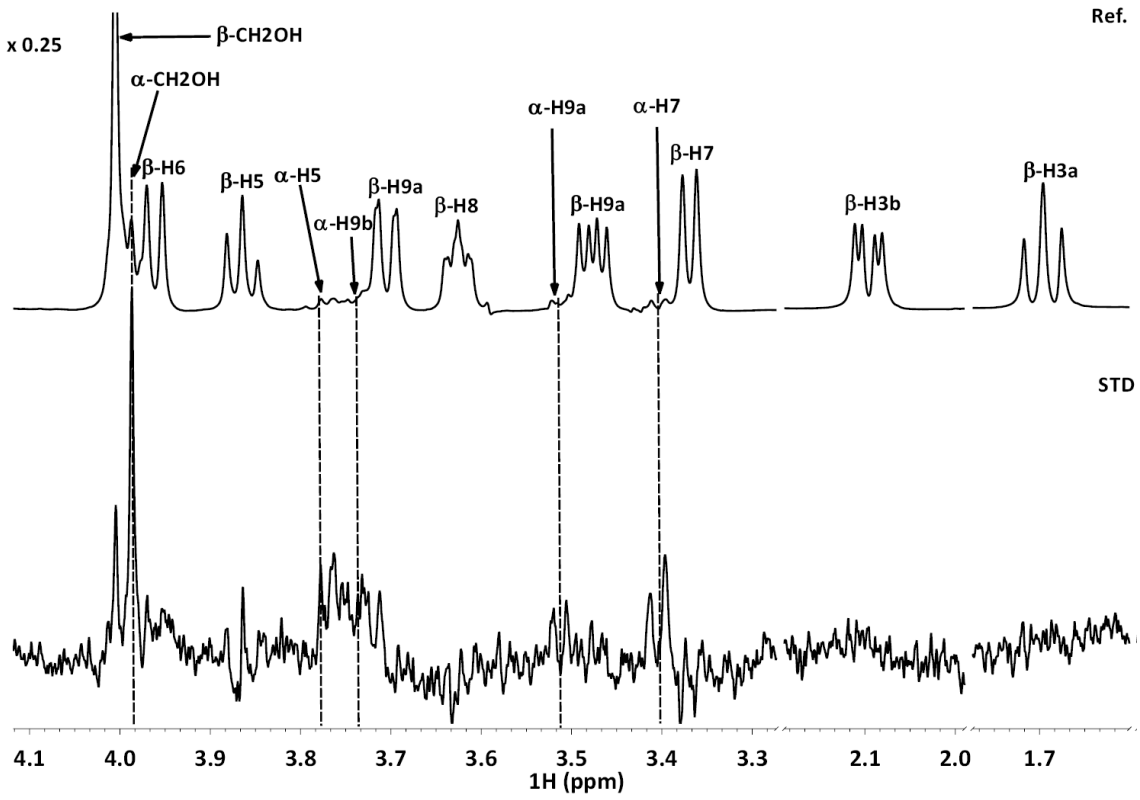


Figure 40. STD (below) and off-resonance (on top) NMR spectra of the interaction of Neu5Gc with the anti-sialic 2 antibody. The temperature was 298 K and a 2 seconds saturation time at the aromatic region (δ 7 ppm) was employed..

	Absolute STD intensity (%)	Relative STD intensity (%)
α-H3a	4.0	63
α-H3b	4.3	68
α-H5	6.2	98
α-H7	5.6	89
α–H9a	2.9	46
α–H9b	6.3	100
α- CH_2OH	2.4	38
β-H5	0.3	4
β–H6	0.3	4
β- CH_2OH	0.2	3

Table 11. The STD intensities estimated for Neu5Gc in the presence anti-sialic 2 antibody.

The epitope mapping analysis (figure 41) revealed similar results to those obtained for anti-sialic **1**. According to the obtained STD intensities, Neu5Gc shows a similar binding epitope versus both anti-sialic antibodies with a similar affinity range.

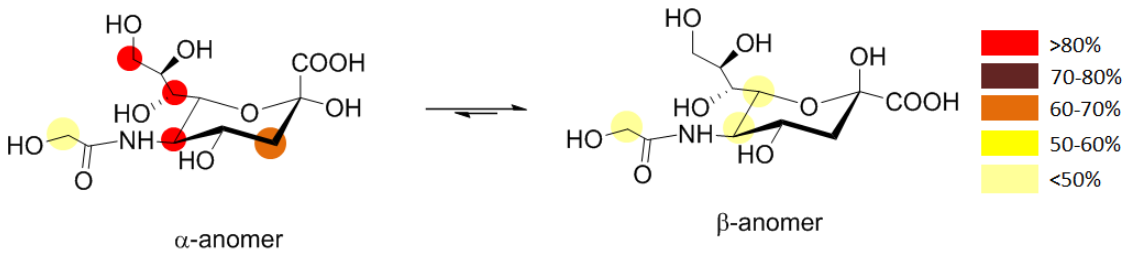


Figure 41. Epitope mapping analysis of Neu5Gc in the presence of the anti-sialic 2 antibody.

Again, no STD was observed for the Neu5Ac analogue in the presence of the anti-sialic **2** antibody under the same experimental conditions employed for the Neu5Gc analogue (figure 42). Both antibodies behave in a similar manner versus the two sialic acid moieties (figures 43 and 44).

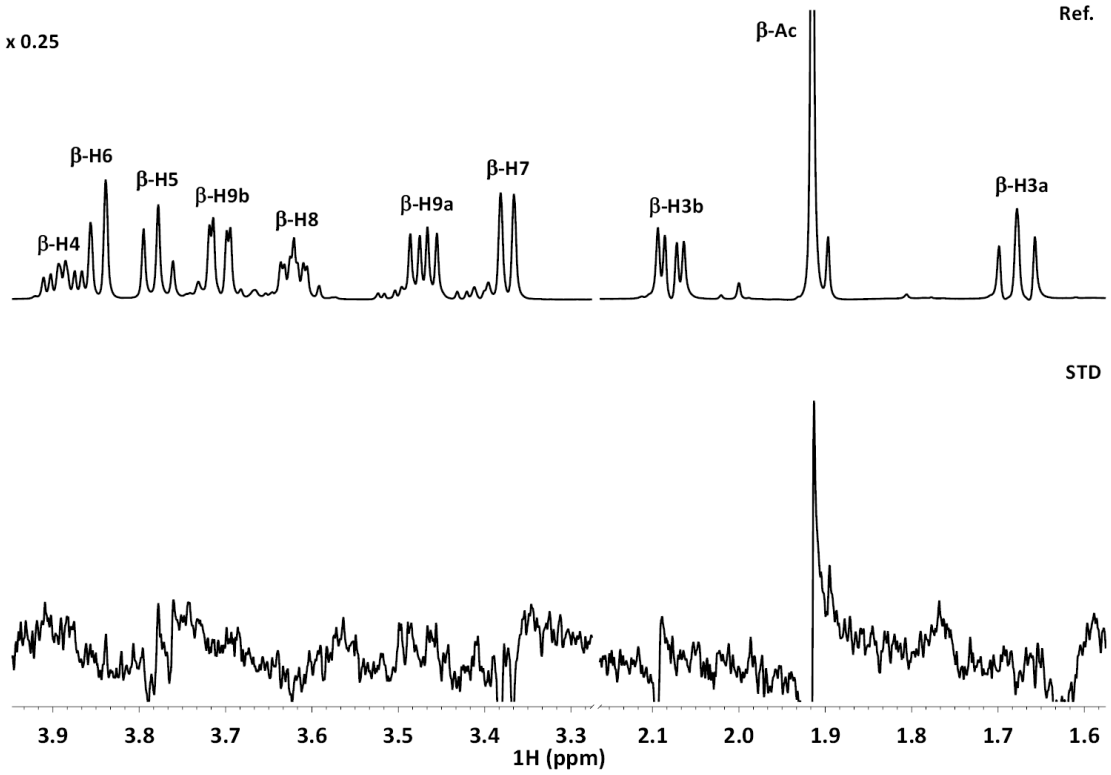


Figure 42. No STD intensities (bottom) are observed for Neu5Ac in the presence of the anti-sialic 2 antibody. The off-resonance spectrum is shown in the top panel.

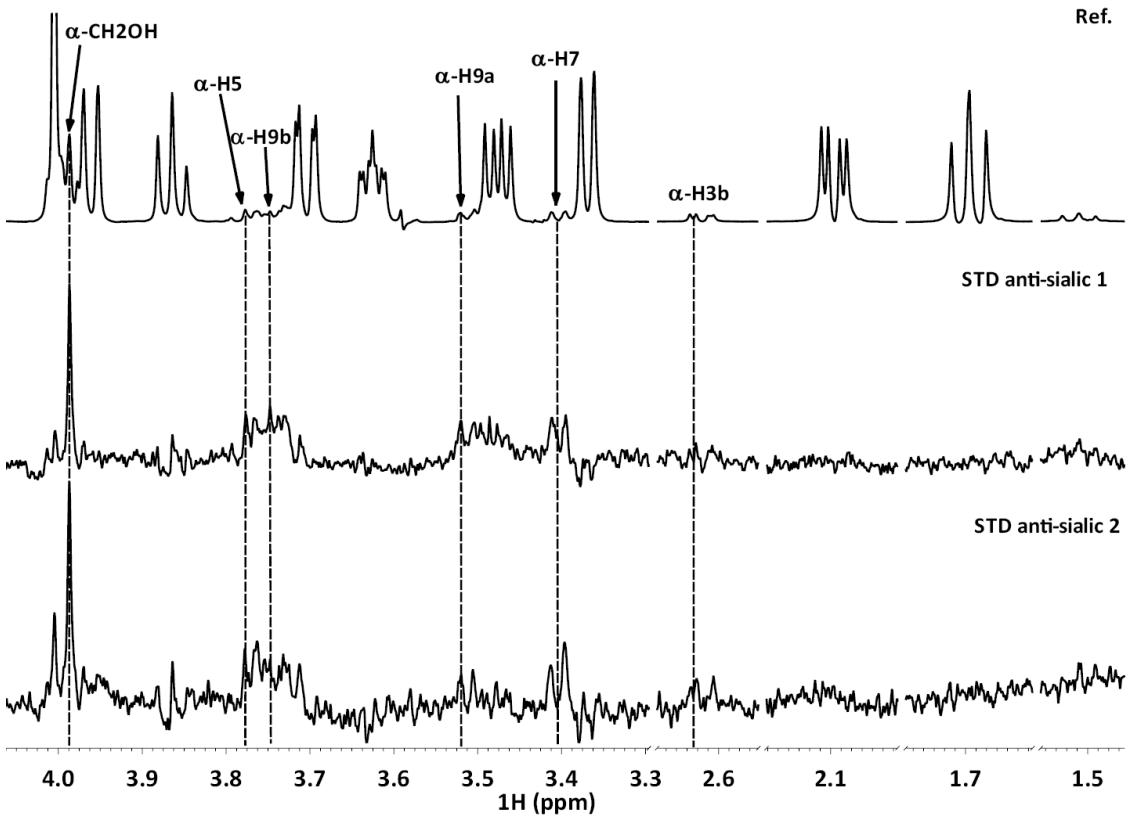
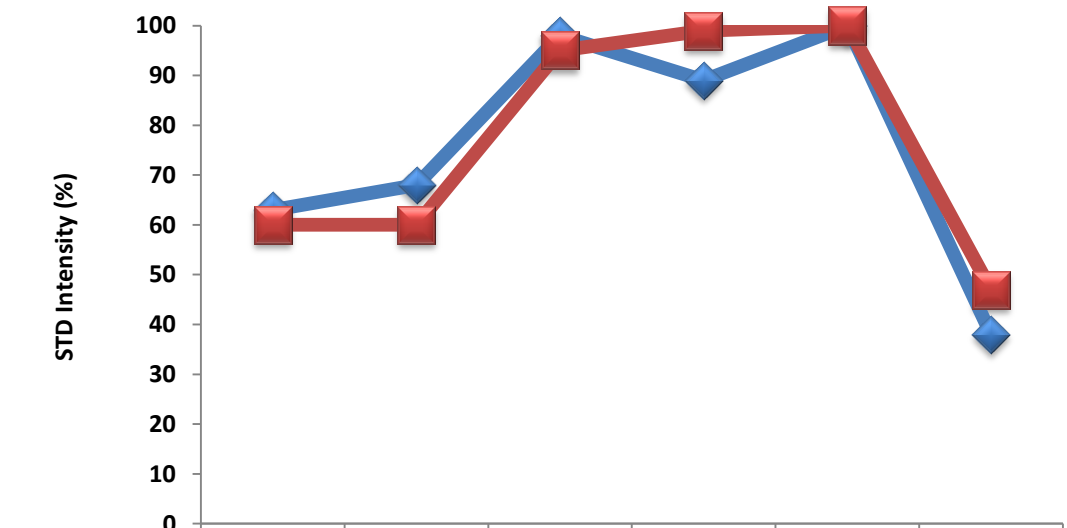


Figure 43. The comparison of the STD-NMR experiments for Neu5Gc in the presence of anti-sialic 1 (central spectrum) and anti-sialic 2 (bottom). The off-resonance spectrum in the presence of anti-sialic 1 is shown at the top.

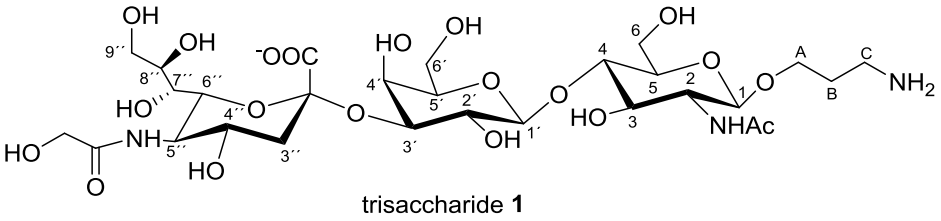


	α-H3a	α-H3b	α-H5	α-H7	α-H9	α-CH2OH
anti-sialic 2	63	68	98	89	100	38
anti-sialic 1	60	60	95	99	100	47

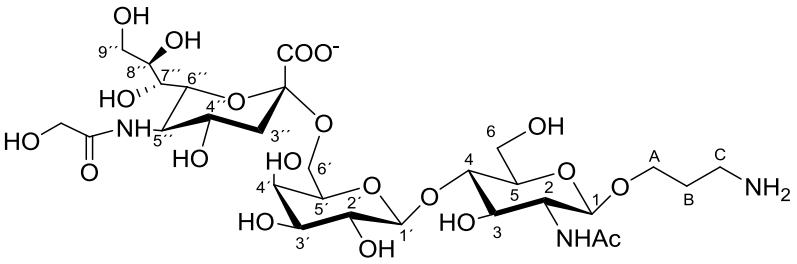
Figure 44. The relative STD intensities of α-Neu5Gc acid in the presence of anti-sialic 1 (grey) and 2 (black) antibodies.

In summary, both monoclonal antibodies display similar affinities and selectivities towards α-Neu5Gc while the Neu5Ac is not recognized either in the α or β anomers.

These simple sialic acid monosaccharides are not directly attached to the cell surfaces but they form part of glycans. Thus, a N-acetyl lactosamine moiety was glycosylated with α-Neu5Gc moiety to yield the naturally occurring Neu5Gcα(2-3)N-acetyl lactosamine and Neu5Gcα(2-6)N-acetyl lactosamine trisaccharide models (figure 45) that were employed to test the monoclonal antibody anti-sialic **1** and **2**, following the same NMR-based methodology.



trisaccharide **1**



trisaccharide **2**

Figure 45. The Neu5Gcα(2-3)N-acetyl lactosamine (1) and Neu5Gcα(2-6)N-acetyl lactosamine (2) trisaccharides.

The ^{1}H NMR spectrum of **1** showed a large overlapping of the resonance signals between δ 3.6 ppm and 3.7 ppm, precluding the non-ambiguous identification of the signals in the STD spectrum.

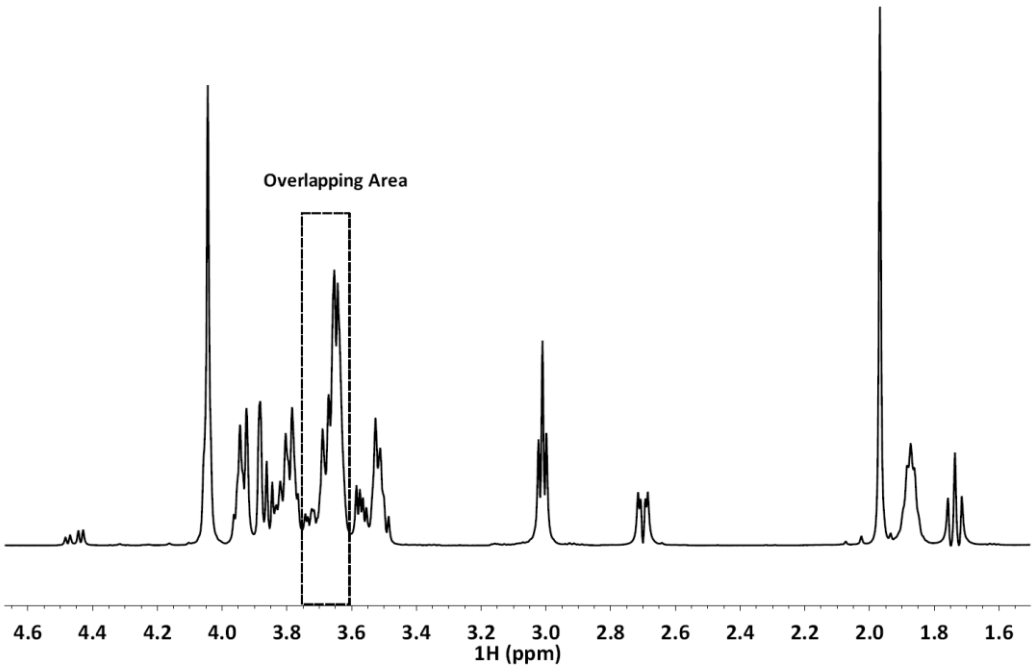


Figure 46. The ^{1}H NMR spectrum (600 MHz) of trisaccharide 1.

Nevertheless, the STD spectrum (figure 47) obtained for trisaccharide **1** in the presence of the anti-sialic **1** antibody showed STD signals for CH_2OH, H6´´, H5´´, and H7´´, indicating that the interaction takes place through the terminal Neu5Gc (figure 48).

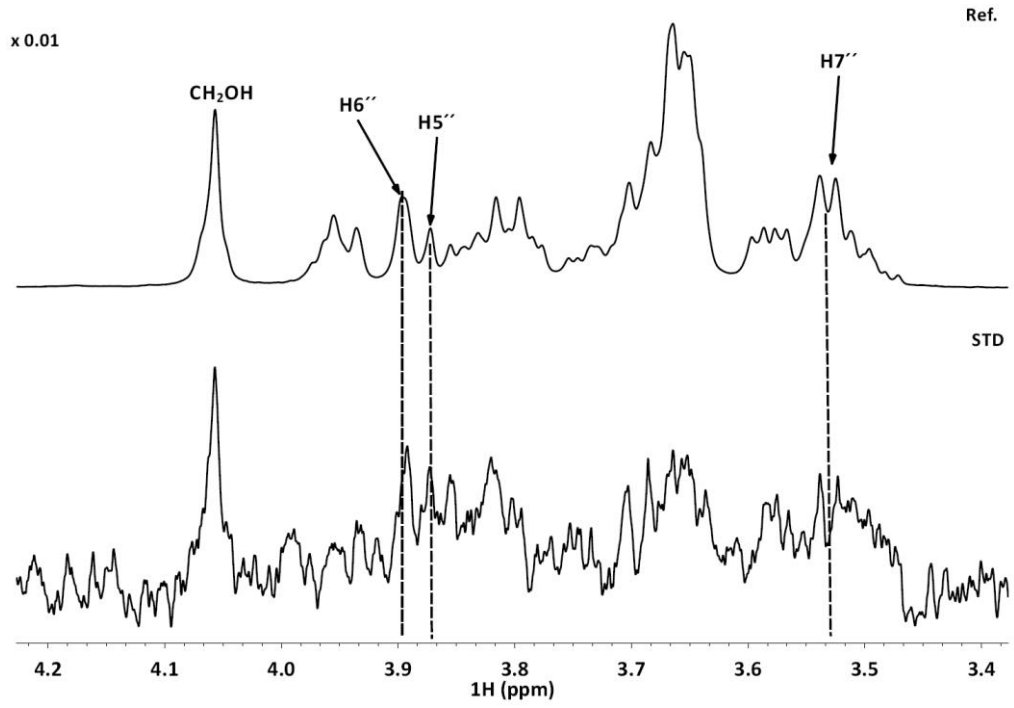


Figure 47. STD (below) and off-resonance (on top) NMR spectra of the interaction of trisaccharide 1 with the anti-sialic 1 antibody. The temperature was 298 K and a 2 seconds saturation time at the aromatic region (δ 7 ppm) was employed.

	Absolute STD intensity (%)	Relative STD intensity (%)
H5´´	2.4	100
H6´´	2.0	81
H7´´	1.5	61
CH_2OH	1.6	68

Table 12. The STD intensities estimated for trisaccharide 1 in the presence of the anti-sialic 1 antibody.

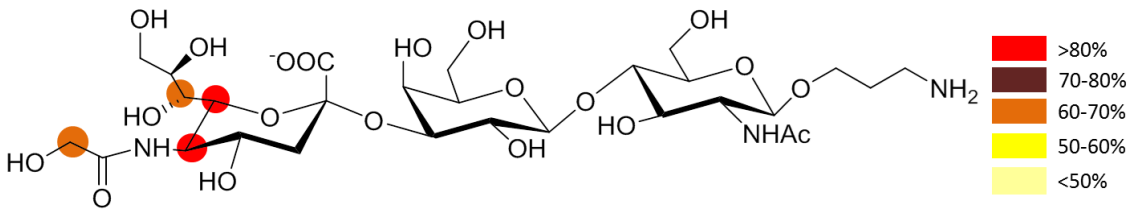


Figure 48. Epitope mapping analysis of trisaccharide 1 in the presence of the anti-sialic 1 antibody.

Then, the binding of trisaccharide **2** towards the anti-sialic **1** antibody was examined under similar experimental conditions. A better signal dispersion was observed in the ^{1}H NMR spectrum of **2**, which largely facilitated the accurate identification and quantification of the STD signals (figure 49). Different sialic acid signals received the highest saturation, but also STD signals from the Gal moiety and even the acetyl group from the GlcNAc residue were evident. Thus, the recognition of **2** by the anti-sialic **1** antibody involves an extended epitope (figure 50).

The chemical natural of the two disaccharides provides major differences in their 3D shapes. Trisaccharide **1**, adopts a rather linear conformation while its analogue **2** displays a "hook-like" geometry. This difference may explain the observed differences in the interaction epitopes: in **2**, once the terminal Neu5Gc is recognized at the binding site of the antibody, the different orientation of the N-acetyl lactosamine moiety provides additional contacts between the Gal and NAc moieties and the antibody.

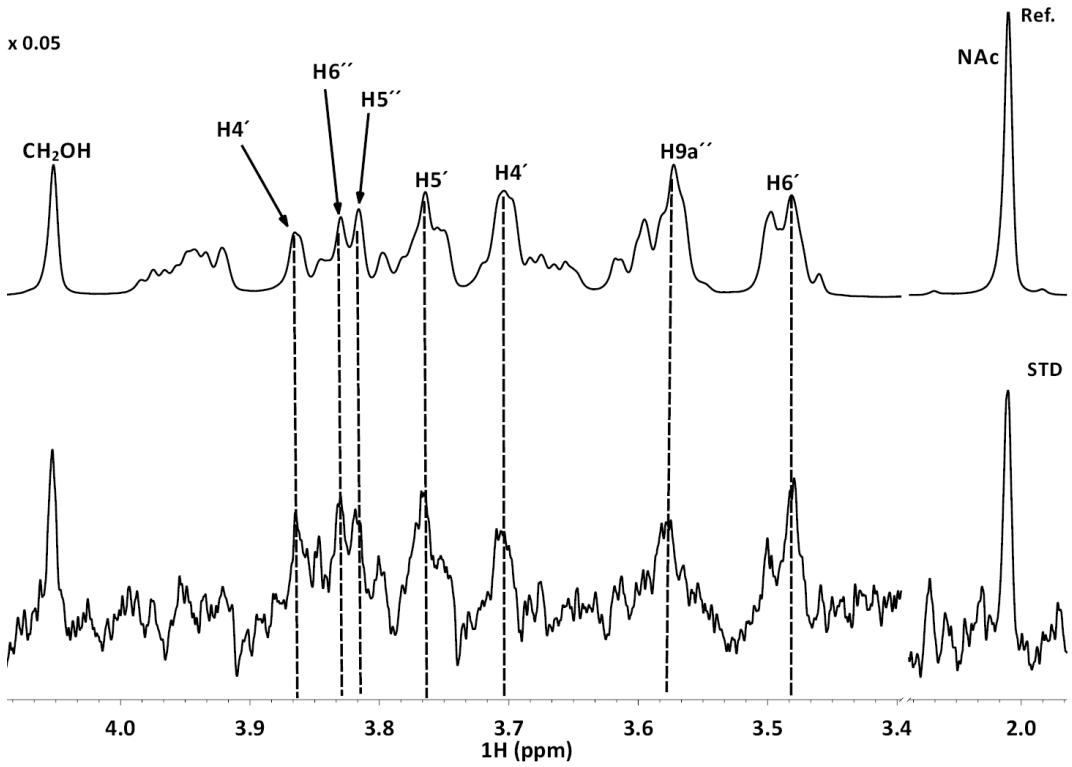


Figure 49. STD (below) and off-resonance (on top) NMR spectra of the interaction of trisaccharide 2 with the anti-sialic 1 antibody. The temperature was 298 K and a 2 seconds saturation time at the aromatic region (δ 7 ppm) was employed.

	Absolute STD intensity (%)	Relative STD intensity (%)
H4'	1.3	100
H5'	1.0	73
H6'	0.9	69
H4´´	0.7	56
H5´´	1.0	77
H6´´	1.2	90
H9a´´	0.9	65
CH_2OH	1.0	72
NAc	0.6	43

Table 13. The STD intensities estimated for trisaccharide 2 in the presence of the anti-sialic 1 antibody.

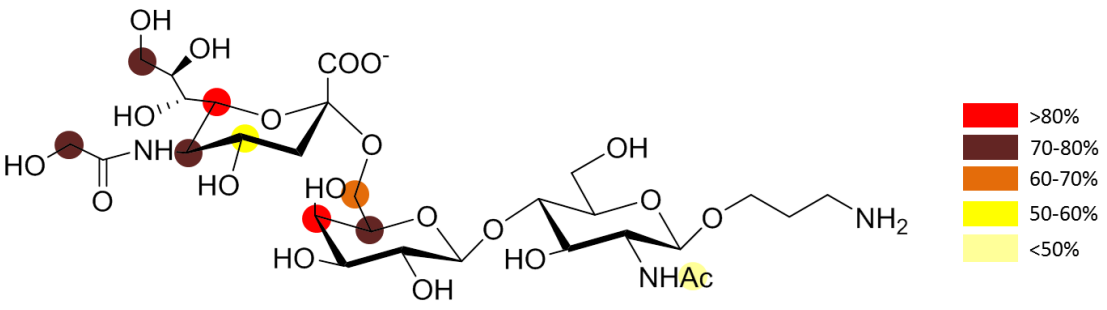


Figure 50. Epitope mapping analysis of trisaccharide 2 in the presence of the anti-sialic 1 antibody.

The same procedure was followed to address the study of the interaction of trisaccharides **1** and **2** with the anti-sialic **2** antibody.

Again, the obtained STD spectrum (figure 51) for trisccharide **1** in the presence of anti-sialic **2** showed clear STD intensities that permitted identifying the CH_2OH group of the sialic acid moiety as a major epitope. Other protons from the Neu5Gc residue were also identified, permitting the description of a similar epitope (figure 52) to that described above for the anti-sialic **1** recognition.

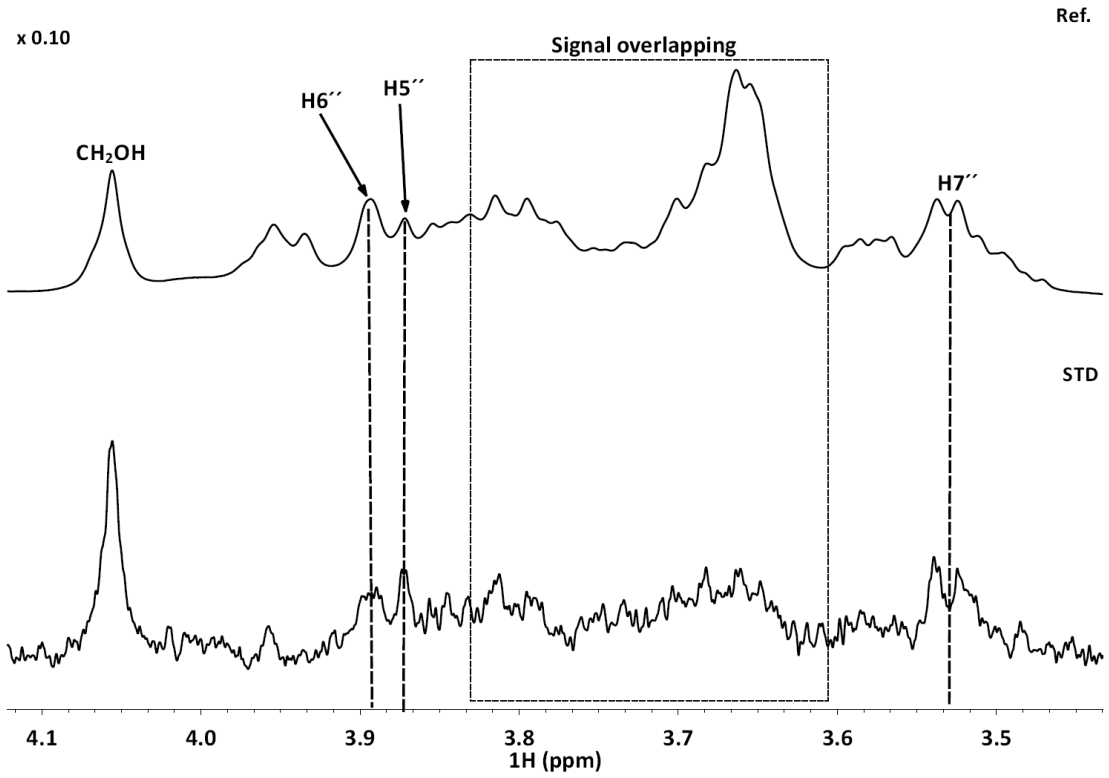


Figure 51. STD (below) and off-resonance (on top) NMR spectra of the interaction of trisaccharide 1 with the anti-sialic 2 antibody. The temperature was 298 K and a 2 seconds saturation time at the aromatic region (δ 7 ppm) was employed.

	Absolute STD intensity (%)	Relative STD intensity (%)
H5´´	14.5	79
H6´´	9.2	51
H7´´	10.0	55
CH_2OH	18.3	100

Table 14. The STD intensities estimated for trisaccharide 1 in the presence of the anti-sialic 2 antibody.

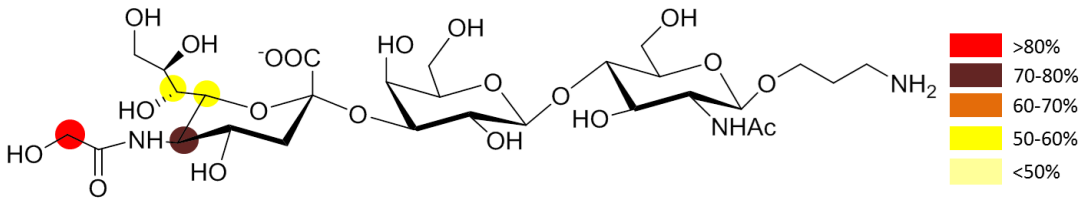


Figure 52. Epitope mapping analysis of trisaccharide 1 in the presence of the anti-sialic 2 antibody.

The analysis of the STD data for trisaccharide **2** in the presence of anti-sialic **2**, (figure 53) showed that the terminal Neu5Gc received the highest saturation and that also the GlcNAc acetyl group was part of the epitope (table 15). The overlapping problem precluded a more defined epitope mapping (figure 54). Thus, the existence of an extended epitope comprising the three saccharide units could not be discarded.

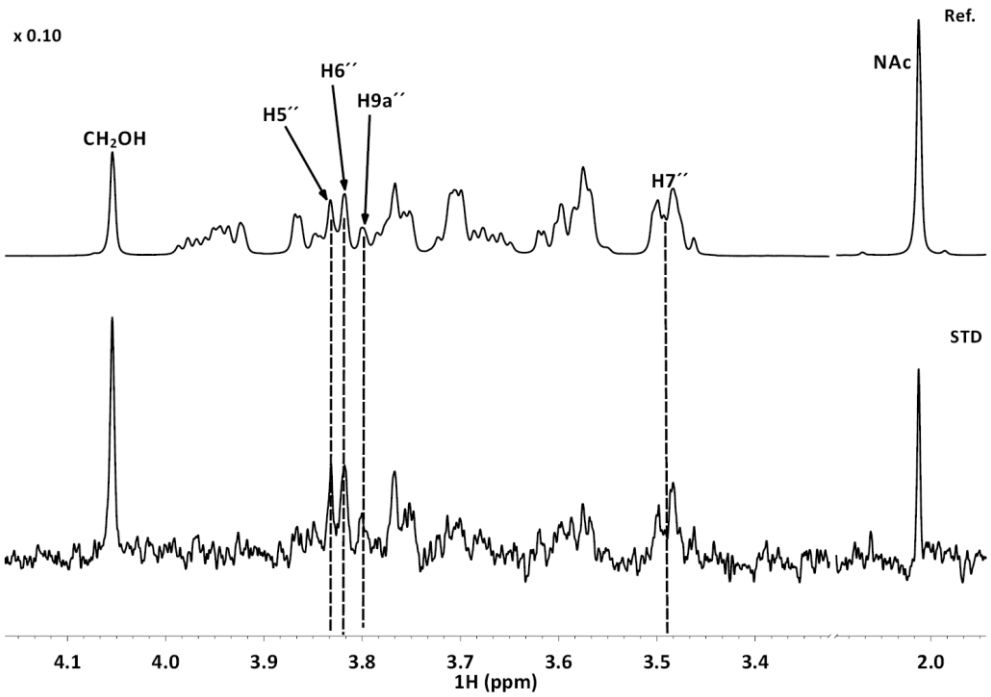


Figure 53. STD (below) and off-resonance (on top) NMR spectra of the interaction of trisaccharide 2 with the anti-sialic 2 antibody. The temperature was 298 K and a 2 seconds saturation time at the aromatic region (δ 7 ppm) was employed.

	Absolute STD intensity (%)	Relative STD intensity (%)
H5´´	1.2	78
H6´´	1.0	68
H7´´	0.8	52
H9a´´	1.2	80
CH_2OH	1.5	100
NHAc	0.5	35

Table 15. The STD intensities estimated for trisaccharide 2 in the presence of the anti-sialic 2 antibody.

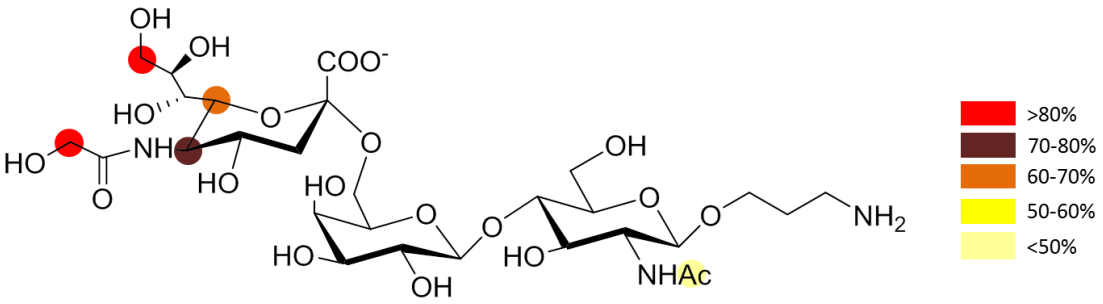


Figure 54. Epitope mapping analysis of trisaccharide 2 in the presence of the anti-sialic 2 antibody

3.4.4. Conclusions

Therefore, both monoclonal anti-sialic antibodies interact with the two α-Neu5Gc-containing trisaccharides. In both cases, the major epitope involves the α-Neu5Gc moiety. For the α-Neu5Gc(2-3) analogue, the antibodies only bind the sialic acid moiety. However, an extended epitope of the Neu5Gc(2-6) analogue, involving the different monosaccharide residues, is recognized by the antibodies. The different 3D geometries of the two molecules are at the heart of the observed molecular recognition behavior.

3.4.5. Experimental

A set of standard 1D and 2D NMR experiments (^{1}H, TOCSY, ^{1}H-^{13}CHSQC, NOESY, ROESY) were performed in a 500 MHz Bruker Avance spectrometer at 298 K to assign the ^{1}H and ^{13}C NMR resonances of the α-gal epitope disaccharide. For the assignment of trisaccharides **1** and **2**, Neu5Ac and Neu5Gc, the experiments were carried out in D_2O and PBS pH 7.3 at 298 K using a 600 MHz Bruker Avance spectrometer equipped with a triple-channel cryoprobe.

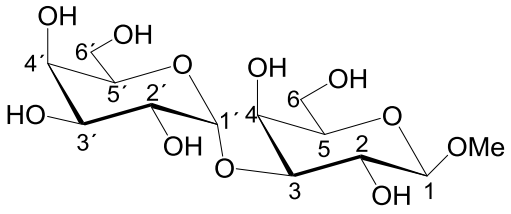
STD-NMR experiments were acquired in the 600 MHz Bruker Avance Spectrometer at 298 K. The saturation frequency was set up at δ 0.8 ppm (aliphatic region) for the anti-gal A4, A11 and G12 antibodies and at δ 7 ppm (aromatic region) for the anti-sialic antibodies. The saturation time was 2 seconds. All the STD-NMR experiments were collected without spin lock. The STD background of the target antibody was deduced by acquiring a blank STD experiment for the pure antibody without ligand and subtracted from the genuine STD in the presence of the corresponding ligand.

For the case of anti-gal A4, A11 and G12 samples, the concentration of the antibody was 10 μM and 900 μM ligand (molar ratio 1:90). The samples contained deuterated PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4; 1.7 mM KH_2PO_4, pH 7.3).

Similar conditions were employed for the preparation of the samples containing anti-sialic samples. For the anti-sialic **1**, the sample concentration of the antibody was 20 μM (molar ratio 1:45) in deuterated PBS buffer at pH 7.3. For the anti-sialic **2** in the presence of trisaccharide **1** spectra, the concentration of the antibody was 20 μM (molar ratio 1:90) in deuterated PBS buffer at pH 7.3.For the anti-sialic **2** in the presence of trisaccharide **2**, the concentration of the antibody was 10 μM (molar ratio 1:90) in deuterated PBS buffer at pH 7.3.

a. α-gal epitope assignment

Standard NMR experiments were acquired on a 5 mM sample of the α-Gal epitope in D_2O. Table 16 shows the assignment of the NMR resonance signals.

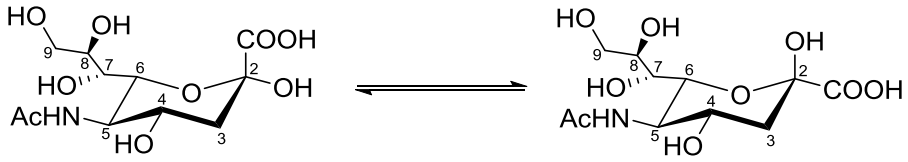


	^{1}H-Chemical Shift (ppm)	Multiplicity (*J* value)	^{13}C- Chemical Shift (ppm)
H1	4.31	d, *J*=7.9 Hz	103.5
H2	3.56	dd, *J*=9.9, 7.9 Hz	69.1
H3	3.70	m	77.3
H4	4.11	m	70.7
H5	3.62	dd, *J*=7.9, 4.34Hz	74.9
H6	3.74	m	61.1
H1'	5.08	d, *J*=3.8 Hz	95.2
H2'	3.79	dd, *J*=10.3, 3.9 Hz	68.2
H3'	3.88	dd, *J*=10.3, 3.3 Hz	69.2
H4'	3.94	d, *J*=3.2 Hz	69.1
H5'	4.11	m	64.7
H6'	3.66	m	60.8
OMe	3.51	s	56.9

Table 16. NMR chemical shifts (δ, ppm) and coupling constants (Hz) for for the α-gal disaccharide epitope.

b. Neu5Ac assignment

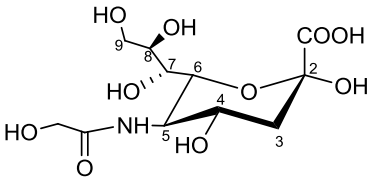
NMR experiments were acquired on 3mM samples of the ligands in D_2O. Tables 17, 18 and 19 show the assignment of the corresponding NMR signals of Neu5Nac and Neu5Gc. Two different set of signals α and β were observed. Their assignments are given in different tables.



	^{1}H-Chemical Shift (ppm)	Multiplicity (*J* value)	^{13}C- Chemical Shift (ppm)
β–H3a	1.72	t, *J* = 13.1 Hz	38.8
β–H3b	2.14	dd, *J* = 13.8, 5.2 Hz	38.8
β–H4	3.90	m	66.7
β–H5	3.78	t,*J*=10.1 Hz	51.9
β–H6	3.89	m	70.3
β–H7	3.39	d, *J* = 9.7 Hz	68.2
β–H8	3.60	m	70.0
β–H9a	3.46	dd, *J* = 12.2, 6.6 Hz	63.1
β–H9b	3.69	dd, *J* = 12.2, 2.3 Hz	63.1
β–Ac	1.90	s	22.1

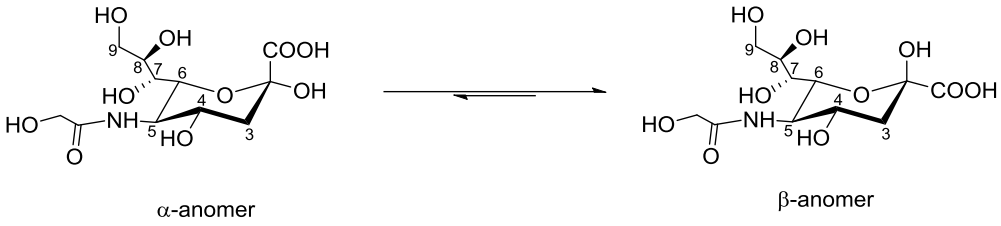
Table 17. NMR chemical shifts (δ, ppm) and coupling constants (Hz) for N-Acetyl Neuraminic acid (Neu5Ac).

c. Neu5Gc assignment



	^{1}H-Chemical Shift (ppm)	Multiplicity (*J* value)	^{13}C- Chemical Shift (ppm)
β–H3	1.73	t, *J* = 13.0 Hz	39.0
β–H3'	2.14	dd, *J* = 13.8, 5.2 Hz	39.0
β–H4	4.00	m	66.5
β–H5	3.86	t, *J*=10.3 Hz	51.7
β–H6	3.99	m	69.9
β–H7	3.38	d, *J* = 9.7 Hz	68.1
β–H8	3.60	m	70.1
β–H9	3.46	dd, *J* = 12.1, 6.5 Hz	63.0
β–H9'	3.69	dd, *J* = 12.5, 2.1 Hz	63.0
β–CH_2OH	3.99	s	60.8

Table 18. NMR chemical shifts (δ, ppm) and coupling constants (Hz) for for N-Glycolyl Neuraminic acid (Neu5Gc).

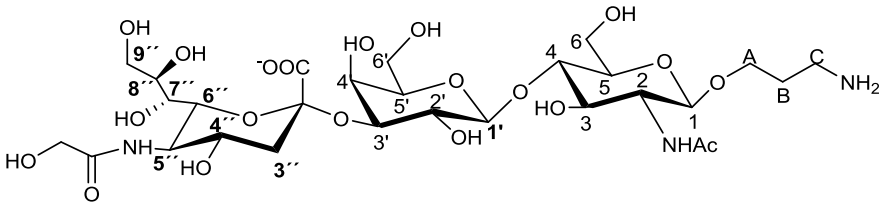


	^{1}H-Chemical Shift (ppm)	Multiplicity (*J* value)	^{13}C- Chemical Shift (ppm)
α –H3	2.62	t, *J* = 12.4 Hz	40.6
α–H3'	1.51	dd, *J* = 12.5, 4.4 Hz	40.6
α–H4	3.51	m	72.2
α –H5	3.76	m	68.2
α –H6	3.87	m	70.0
α –H7	3.37	m	67.9
α –H8	3.73	m	71.4
α –H9a	3.49	m	62.5
α –H9b	3.72	m	62.5
α–CH2OH	3.99	s	60.8
β –H3a	1.69	m	39.3
β–H3b	2.09	m	39.3
β–H4	3.99	m	66.8
β –H5	3.87	m	51.8
β –H6	3.96	m	69.8
β –H7	3.37	m	68.4
β –H8	3.63	m	70.2
β –H9a	3.48	m	63.2
β –H9b	3.71	m	63.2
β–CH_2OH	4.00	m	60.8

Table 19. NMR chemical shifts (δ, ppm) and coupling constants (Hz) for Neu5Gc acid.

d. Trisaccharides assignment

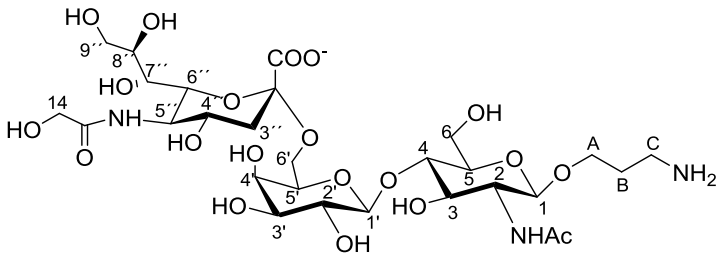
NMR experiments were acquired for 3 mM samples of the ligands in D_2O. Tables 20 and 21 show the NMR parameters for each ligand.



	^{1}H-Chemical Shift (ppm)	Multiplicity (*J* value)	^{13}C- Chemical Shift (ppm)
H1	4.43	d, *J* = 8.2 Hz	101.2
H2	3.68	m	55.1
H3	3.68	m	72.6

H4	3.64	m	78.3
H5	3.51	m	74.7
H6a	3.57	dd, *J*=11.9,6.3 Hz	62.5
H6b	3.79	m	62.5
H1'	4.48	d, *J* = 7.9 Hz	102.5
H2'	3.50	m	69.3
H3'	4.04	m	75.5
H4'	3.64	m	72.1
H5'	3.63	m	75.2
H6a'	3.78	m	59.9
H6b´	3.93	m	59.9
H3a´´	2.70	dd, *J* = 12.6, 4.6 Hz	40.0
H3b´´	1.74	t, *J*=12.1 Hz	40.0
H4´´	3.71	m	68.0
H5´´	3.86	m	51.3
H6´´	3.88	m	67.4
H7´´	3.51	m	68.0
H8´´	3.82	m	71.8
H9a´´	3.65	m	61.0
H9b´´	4.03	m	61.0
CH_2OH	4.04	m	61.0
A	3.64	m	68.0
A´	3.94	m	67.9
B	1.87	m	27.0
C	3.01	t, *J*=6.9 Hz	37.9
Ac	1.96	s	102.2

Table 20. NMR chemical shifts (δ, ppm) and coupling constants (Hz) for trisaccharide 1.



	^{1}H-Chemical Shift (ppm)	Multiplicity (J value)	^{13}C- Chemical Shift (ppm)
H1	4.47	dd, *J*= 5.8, 2.5 Hz	101.0
H2	3.69	m	54.8
H3	3.56	m	74.4
H4	3.56	m	80.6
H5	3.55	m	84.4
H6	3.76	m	60.2
H6b	3.92	m	60.2
H1'	4.38	d, *J*= 7.9 Hz	103.4
H2'	3.46	m	70.7

H3'	3.59	m	72.4
H4'	3.85	m	68.3
H5'	3.75	m	73.7
H6'	3.47	m	63.4
H6b´	3.93	m	63.4
H3a´´	2.61	dd, *J*= 12.4, 4.7 Hz	40.2
H3b´'	1.65	t, *J*=12.2 Hz	40.2
H4´´	3.69	m	68.0
H5´´	3.82	m	51.5
H6´´	3.82	m	71.8
H7´´	3.47	m	68.3
H8´´	3.85	m	78.4
H9a´´	3.57	m	62.6
H9b´´	3.80	m	62.6
CH_2OH	4.04	s	61.0
A	3.65	m	67.9
A´	3.96	m	67.9
B	1.88	dd, *J*=9.2, 4.1 Hz	27.0
C	3.03	t, *J*= 6.9 Hz	37.6
NHAc	1.99	s	102.2

Table 21. NMR chemical shifts (δ, ppm) and coupling constants (Hz) for trisaccharide 2.

3.5. Sialylated glycans and influenza

As mentioned above, sialic acid moieties at the glycocalyx plays a key role human upper respiratory tracts.[94,97,98] Thus, it is exposed to pathogens through the air, becoming involved in common pathologies such as influenza.

Influenza viruses are a family of RNA viruses responsible for flu and are associated with several major human pandemics in the 20th century. One major concern resides in its ability to mutate and then to infect other species. Its genome codifies for 10 proteins. Two of them are located at the surface of the virus particles. The first one, Hemagglutinin (HA) is important for cell adhesion and infection. The second one, Neuraminidase (NA) is the key player for virions release.[99,100] Both proteins are characterized by its ability to bind sialic acid residues. Indeed, the transmission between species of the influenza virus relies on the mutation of only a few amino acids of these proteins, which are strongly involved in the binding to sialic acid. At present, 16 hemagglutinins and 9 neuraminidases subtypes of influenza A have already been described.[101–104] Due to the large variety of combinations and origin diversity of the proteins (avian, human, other mammals), the different influenza viruses have been named by the presence of hemagglutinins from H1 to H16 and of the corresponding neuraminidases, from

N1 to N9. For instance, the H1N1 considers the presence of the hemagglutinin subtype 1 (H1) and the neuraminidase subtype 1 (N1) at the surface of the influenza virus. Moreover, the influenza viruses that infect humans have also been classified, based on their virulence and pathogenicity, in three types A, B and C affecting humans. While human influenza A and B viruses cause seasonal epidemics, influenza C causes only mild illness.[105]

Sialylated glycan receptors at the host cells have been shown to display a key role for influenza viral infection since they serve as anchorage for the influenza virus, via HA recognition, allowing the virus invasion and infection. The easy mutation ability of the HA has triggered the transfection from animal species to humans. The avian and porcine influenza viruses showed specificity towards α2-3 sialic acid linked to galactose (Gal), while many studies have shown that human influenza strains binds α2-6 sialic acid linkages[97,106–115] The single mutation of one or two amino acids at the binding site of the avian influenza strains is enough to switch the specificity from α2-3 sialylated glycans to the α2-6 analogues, thus triggering human infection. Therefore, the full understanding of the HA specificity and binding mode would help to the design of better drugs targeting the pandemic influenza.

3.5.1. Results and discussion

In this work, a small library of five sialylated linear and branched glycans (figure 55, scheme 4) has been synthesized by sequential enzymatic synthesis using a battery of glycosyl transferase enzymes taking advantage of the enzymatic substrate selectivity and specificity, as well as the absence of side products. These factors make the enzymatic approach highly reliable for the synthesis of glycans, avoiding the protecting and deprotecting steps typical of the chemical synthesis of carbohydrates. Next, the recognition of the linear glycans (glycans 1 to 3 figure 55, scheme 4) by different HA strains have been elucidated by using NMR methods from the ligand point of view.

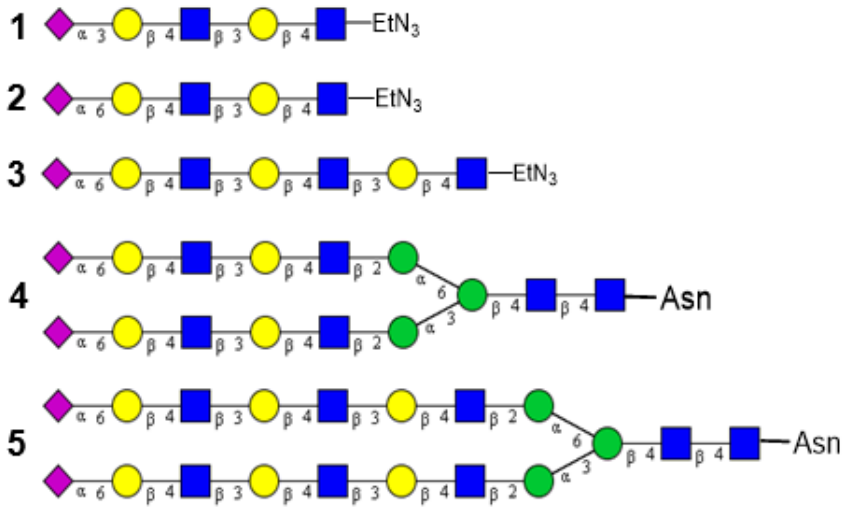
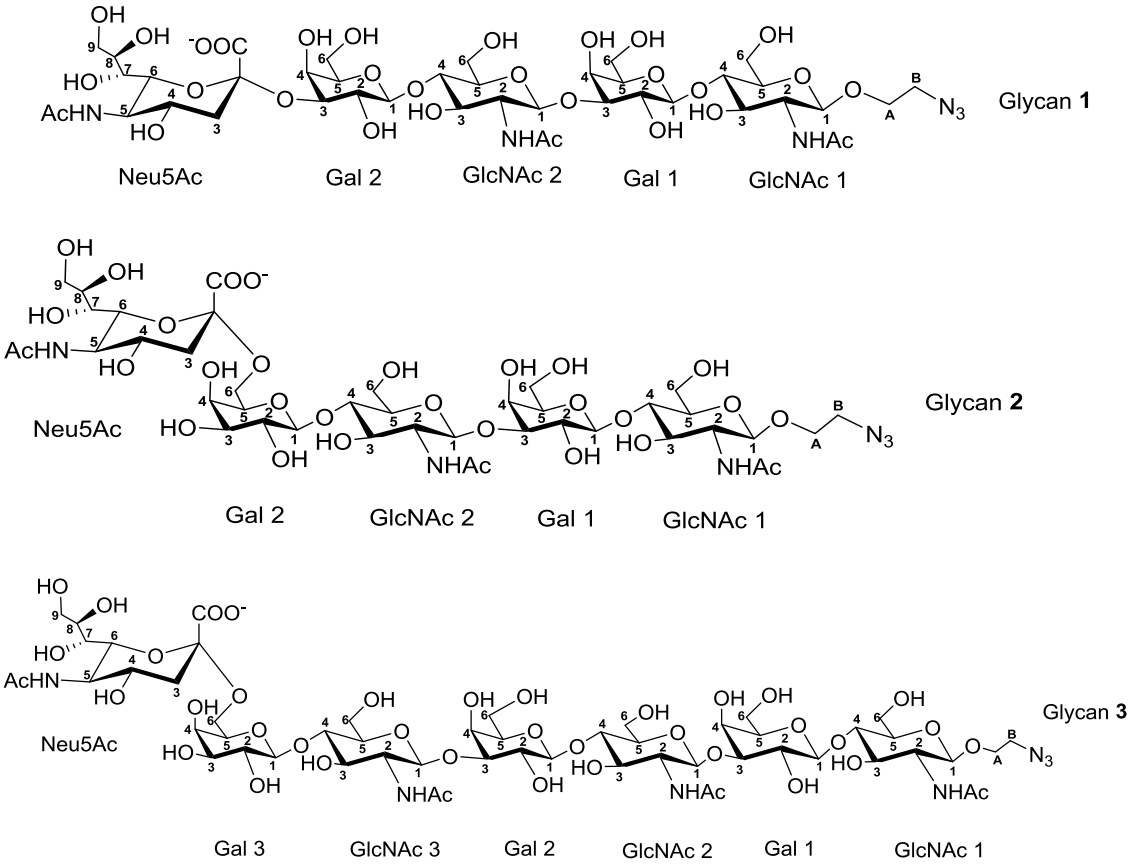


Figure 55. Schematic structures of ligands 1-5.



Scheme 4. Formulae of the linear glycans 1-3 and numbering code of the constituent monosaccharide units.

The synthetic route followed for the synthesis of the linear ligands **1** to **3** is shown in figure 56 From the starting building blocks, successive Gal and GlcNAc extensions were followed, in an alternative manner, to finally extend the oligosaccharide to intermediates **3b** and **1c**. Finally,

the Neu5Ac was added to the terminal Gal with the different α2-3 (avian) or α2-6 (human) linkages.

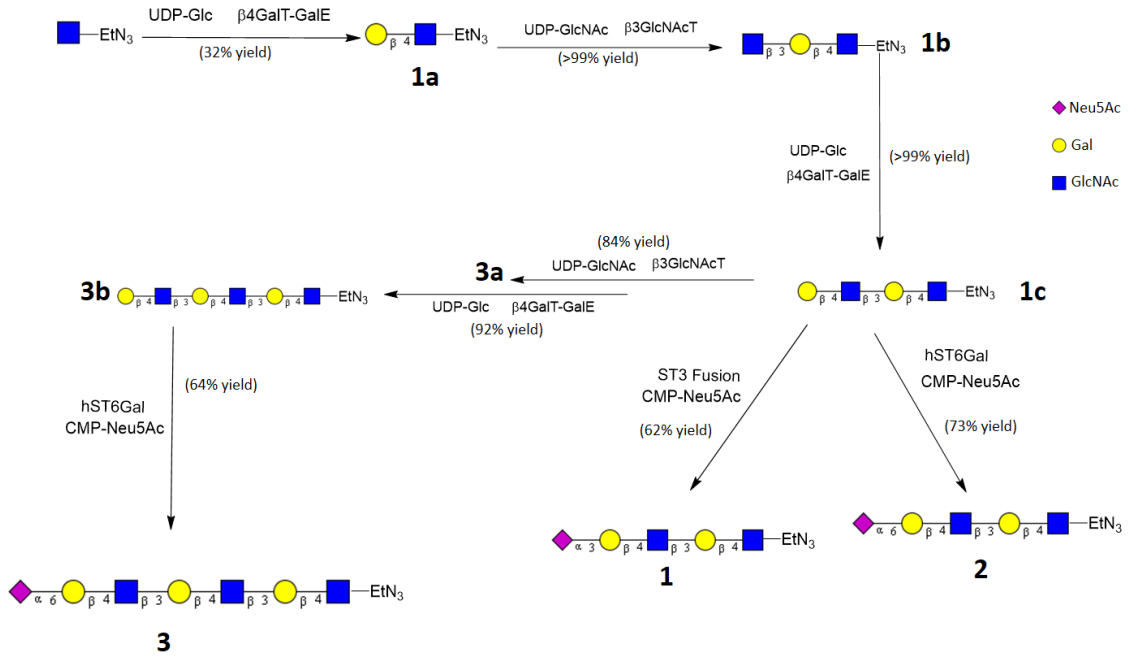


Figure 56. Synthetic routes of glycans 1 to 3.

As reactants, glycosyl donor in their UDP- or CMP- activated carbohydrates are required for the enzymes, which are transferred to the main scaffold by the activity of different transferases. In the case of the Gal extension, the reactant UDP-Glc is first epimerized to Gal and then transferred, by the action of a recombinant enzyme with epimerase and transferase activity (β4GalT-GalE), to form a specific β1-4 linkage with the terminal GlcNAc ethyl azide as starting building block. The next GlcNAc addition is directly added to the growing oligosaccharide chain from its activated UDP- form by the activity of the GlcNAc-transferase, β3GlcNAcT, to form a β1-3 linkage to the Gal unit. These two steps are sequentially repeated once (glycans **1** and **2**) or twice (glycan **3**). In the last steps of the synthesis, for the addition of the Neu5Ac residue, two different sialyl transferase enzymes with the same sialyl donor substrate, CMP-Neu5Ac previously obtained by enzymatic sytnhesis, have been used to address the different linkages, α2-3 (glycan **1**) or α2-6 (glycans **2** and **3**): the ST3 fusion sialyl transferase yields the α2-3 linkage, while the hST6Gal sialyl transferase incorporates the corresponding α2-6 linkage to the terminal Gal. The experimental section contains the detailed description of the conditions of the different reactions.

The synthesis of the biantennae ligands (glycans **4** and **5**) was accomplished using a similar strategy (figure 57), but employing a key branched nonasaccharide starting building block (**4b**) obtained by enzymatic degradation of a complex undecasaccharide (**4a**) isolated from natural sources. This preliminary step was required to remove the terminal neuraminic acid of the starting material (**4a**) with a comercial sialidase (AUS neuraminidase, Sigma Adrich). From **4b**, sequential growing steps were run similarly as for the linear oligosaccharides.

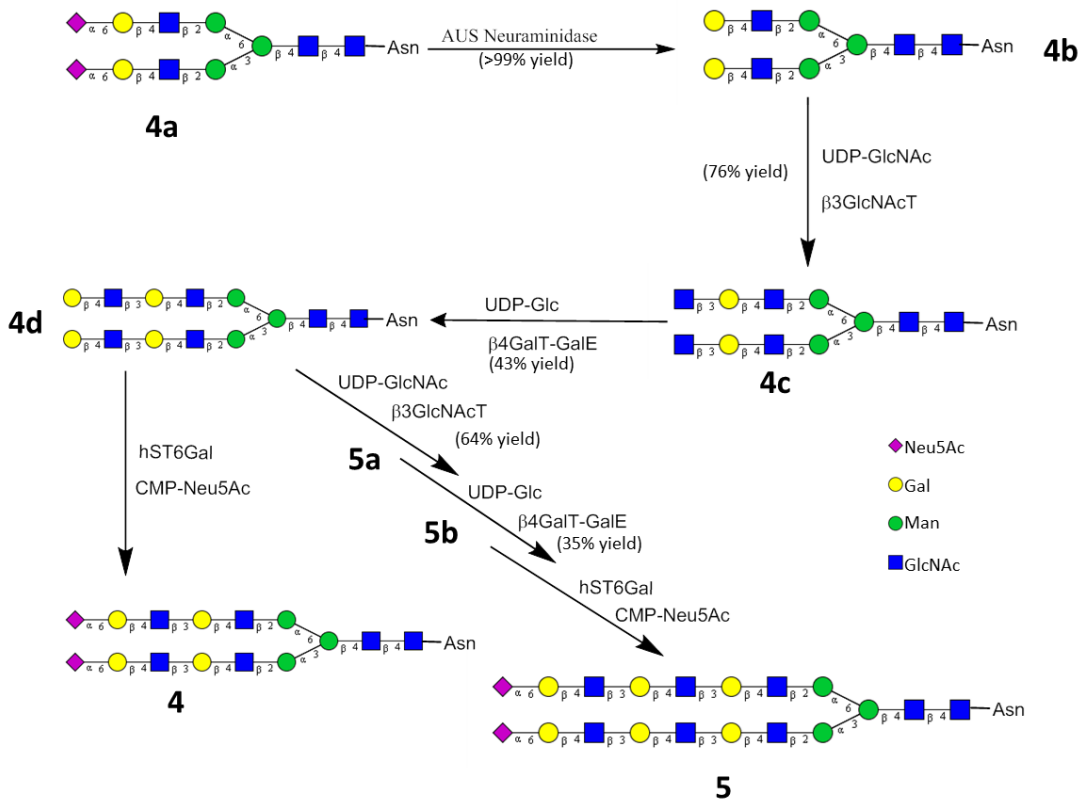


Figure 57. Synthetic routes to the branched ligands 4 and 5.

After the synthesis, a set of standard NMR experiments (TOCSY, NOESY, ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HSQC-TOCSY) were collected for chemical shift assignments (see experimental section). To differentiate the signals corresponding to each N-Acetyl group present in the Neu5Ac and GlcNAc residues a combined set of TOCSY and NOESY experiments acquired in water (10% deuterated) were needed (figure 58). In all cases, the acetyl group of Neu5Ac residue appeared at higher field.

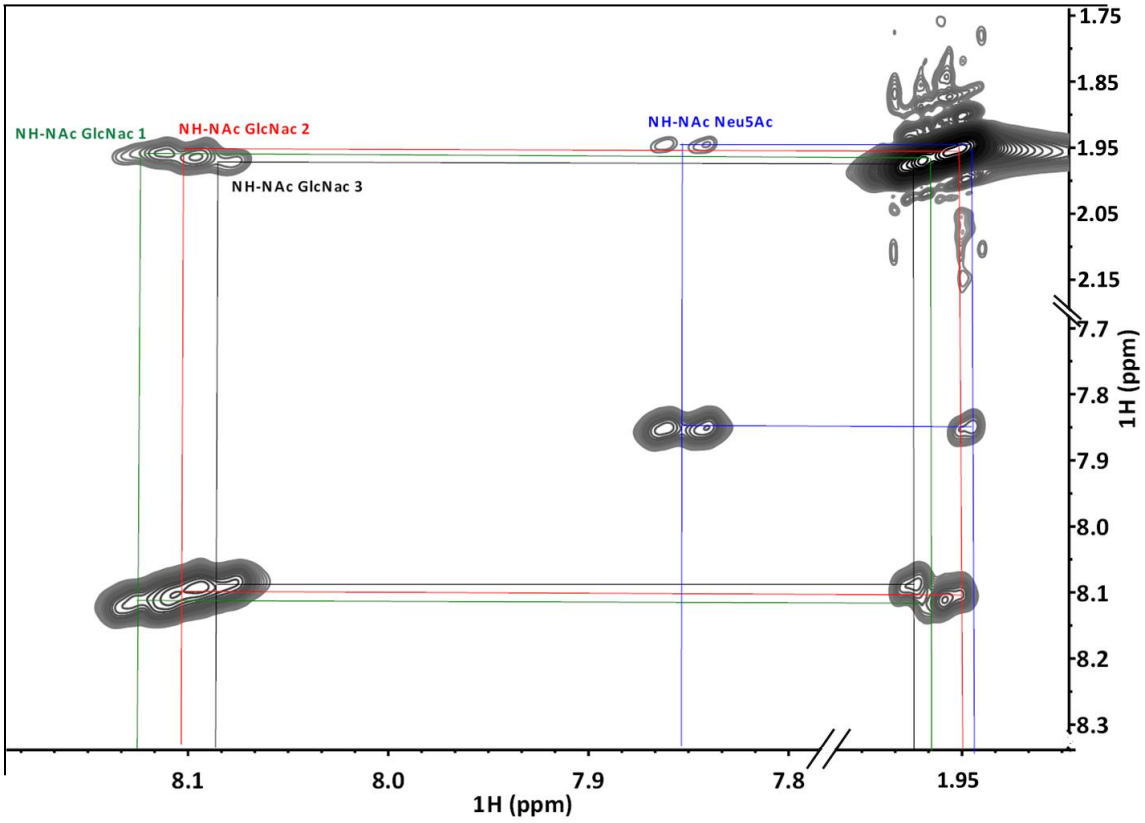


Figure 58. NOESY spectrum of glycan 3 in water (10% deuterated) at 600 MHz equipped with a cryoprobe (298 K). The NH-NAc cross-peaks are annotated.

The glycans were then tested against the influenza A human variants of H1N1, H7N9 and H6N1 strains and to the original avian strains of H7N9 and H6N1. The human strains display affinity towards the Neu5Ac α2-6Gal linkage as evidenced by Nycholat *et al* by using glycoarrays[116] and by Macchi *et al* by using NMR.[117] Due to the large protein size of HA variants (a trimer of 60 KDa per monomer), the use of ^{1}H-^{15}N HSQC titrations to analyze the binding event form the protein's perspective is not suitable for this case. Therefore, we again employed NMR methodology from the ligand´s point of view with STD NMR experiments.

The analysis of the STD spectrum of **2** in the presence of the HA of the human H1N1 strain (figure 59) identified the acetyl group of the Neu5Ac as the most saturated signal, followed by H3a and H3b of Neu5Ac and the acetyl group of GlcNAc 2. Although it was not possible to completely analyze the crowded region from δ 3.5 to 4 ppm due to extensive signal overlapping, interestingly, H4 of Gal 1 and Gal 2 residues showed saturation transfer from the protein, indicating the presence of an extended epitope along the complete oligosaccharide chain.

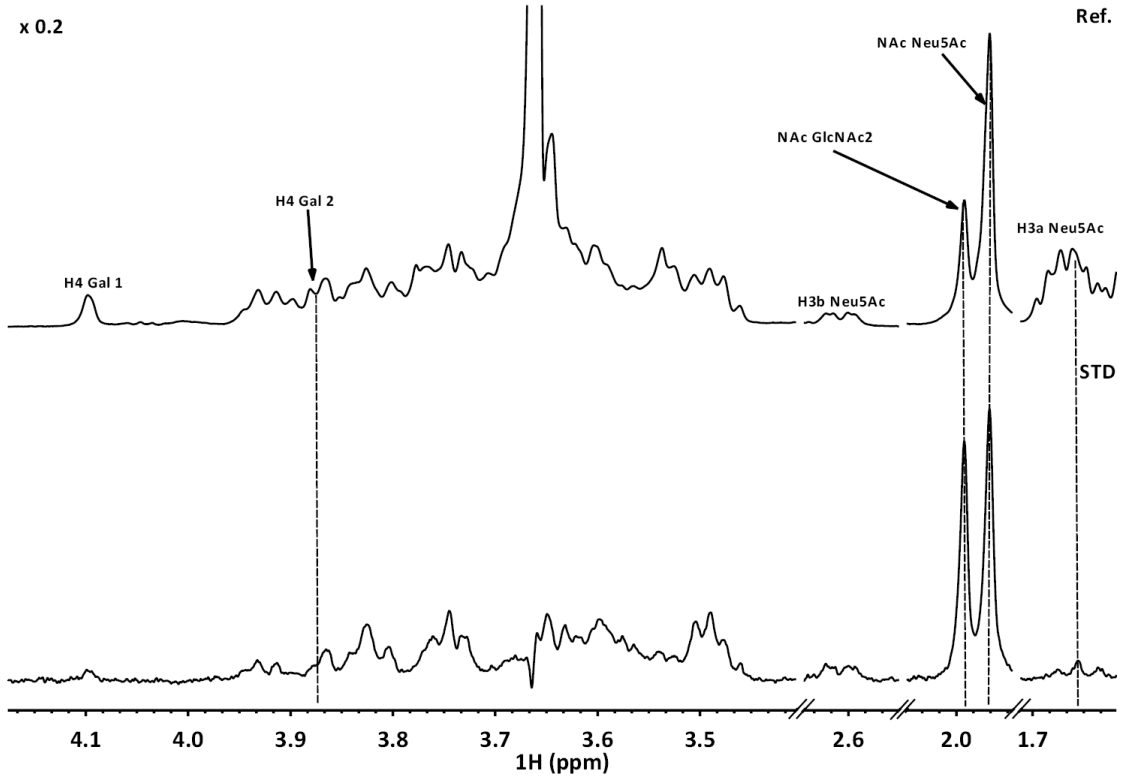


Figure 59. The interaction of glycan 2 with H1N1 strain of HA. The STD spectrum (600 MHz, 298 K) obtained for a 86:1 mixture of glycan 2:H1N1 is shown at the bottom, while the off-reference spectrum is shown at the top. The concentration of the HA was 3.5 μM. The on-resonance frequency for protein saturation was set at the aromatic region (δ 7 ppm) during 2 seconds.

Similar results were obtained from the STD experiments performed for the complex of the HA of the H1N1 variant and compound **3** (figure 60). Similarly, the main contact is located at the Neu5Ac end of the molecule. In fact, the acetyl group of the Neu5Ac receives the highest saturation followed by the H3a and H3b of the Neu5Ac and the acetyl group of GlcNAc 3. The discrimination among the different acetyl groups is remarkable, suggesting that only the acetyl groups from the Neu5Ac and GlcNAc 3 residues are in close contact with the protein. Again, the signals of H4 from Gal 3 and overlapping signal from Gal 1 and Gal 2 residues showed saturation, as expected for the extended epitope observed with **2**, which would involve at least the galactose units 2 and 3.

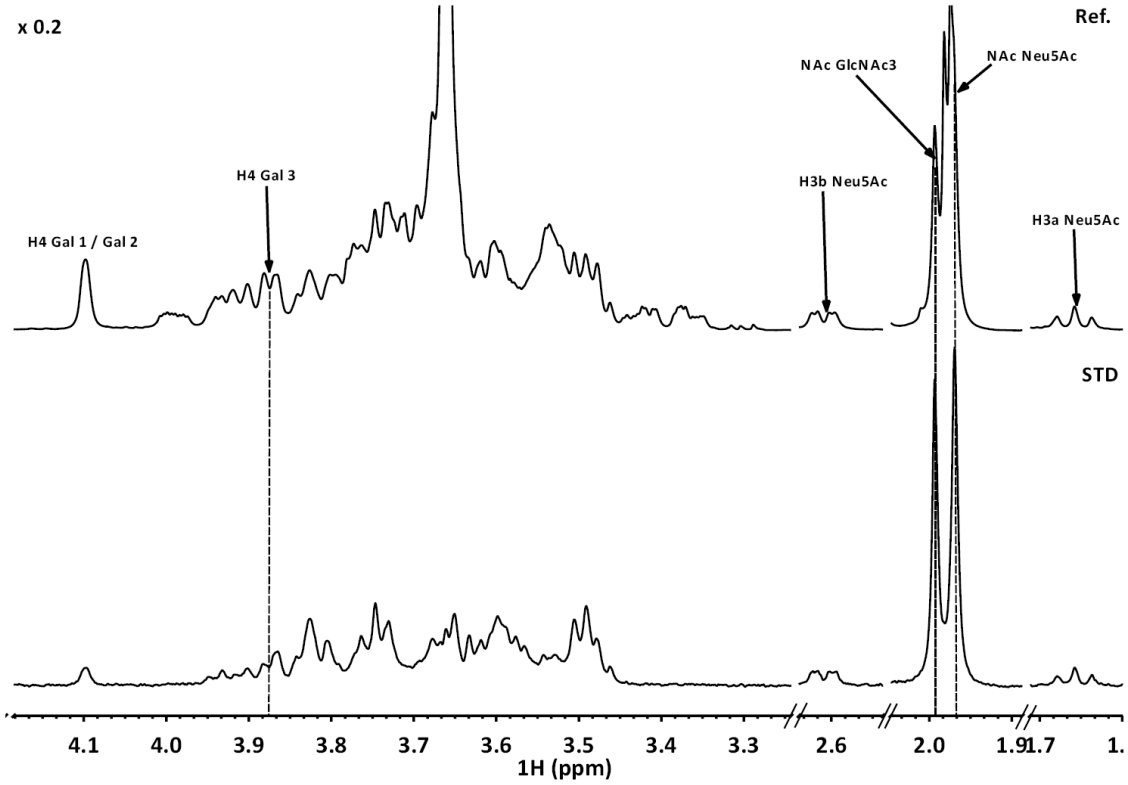


Figure 60. The interaction of glycan 3 with H1N1 strain of HA. The STD spectrum (600 MHz, 298 K) obtained for a 86:1 mixture of glycan 3:H1N1 is shown at the bottom, while the off-reference spectrum is shown at the top. The concentration of the HA was 3.5 μM. The on-resonance frequency for protein saturation was set at the aromatic region (7 ppm) during 2 seconds.

For a better understanding of the binding event, a computational model was generated from the available H1N1 human variant PDB structure (PDB ID 4JTV).[118] Thus, glycan 3 was modeled using the GLYCAM force field.[119] The built model suggests that the binding site of the HA involves loop 130, loop 220, and the alpha helix 190 (left panel of figure 61). Indeed, the experimental STD intensities are in accordance with the built binding mode and suggest the existence of a fully folded umbrella-type topology at the glycosidic bond Neu5Acα(2-6)Gal, similar to that already described for other strains.[117,120–122] The analysis of the obtained geometry permitted to deduce that the HA efficiently binds the glycan by favoring the gauche-trans (gt) conformation for the ω dihedral angle of the terminal Gal 3 moiety (O_{6Gal3}-C'_{6Gal3}-C'_{5Gal3}-O'_{5Gal3} (gauche) and O_{6Gal3}-C'_{6Gal3}-C'_{5Gal3}-C'_{4Gal3} (trans)) and selecting the typical extended geometry for the subsequent β1-3 and β1-4 glycosidic linkages (right panel of figure 61).

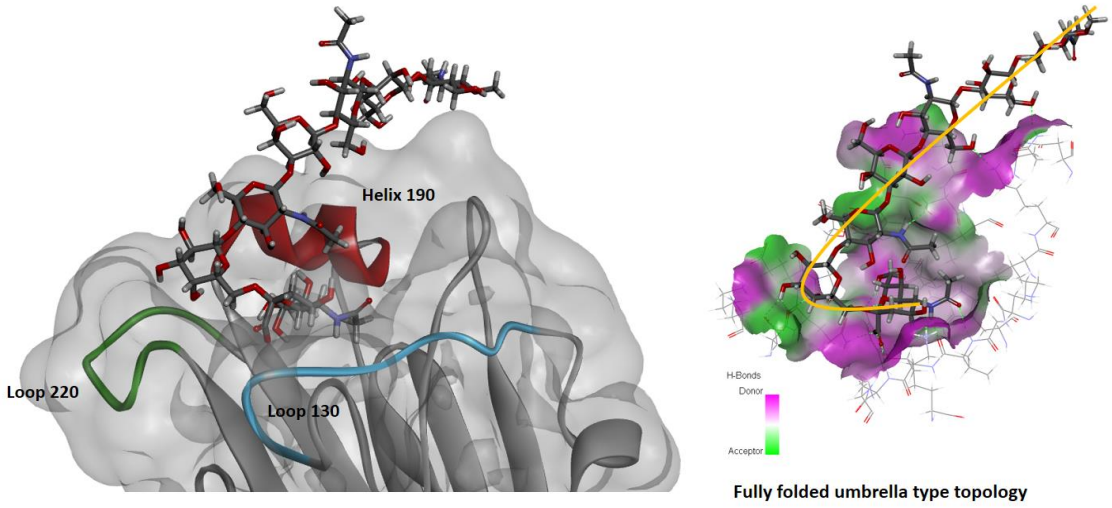


Figure 61. Proposed 3D model for the complex formed by the interaction of glycan 3 with the HA of the H1N1 variant. The binding site is represented, showing the topology of the glycan and the interaction pattern.

As mentioned in the introduction to this section, the importance of single mutations between different strains may drive the change in the specificity of the HA from viruses with different targeted species and, therefore, the study of the binding specificity of mutants display special relevance. Hence, glycan **3** was tested against two different HA influenza A avian strains (H6N1 and H7N9), specific for α(2-3) Neu5Ac-Gal linkages, and against their mutated relatives targeting humans, specific for the α2-6 analogues.

The STD-NMR spectra of the avian wild type (WT) and mutant human variant H6N1 strains in the presence of glycan **3** are shown in figure 62. From the higher efficiency in the saturation transfer, it is evident that the mutant (α2-6 specific) strain shows an enhanced binding for **3** in comparison to that of the WT strain, which is specific for the α2-3 linkage.

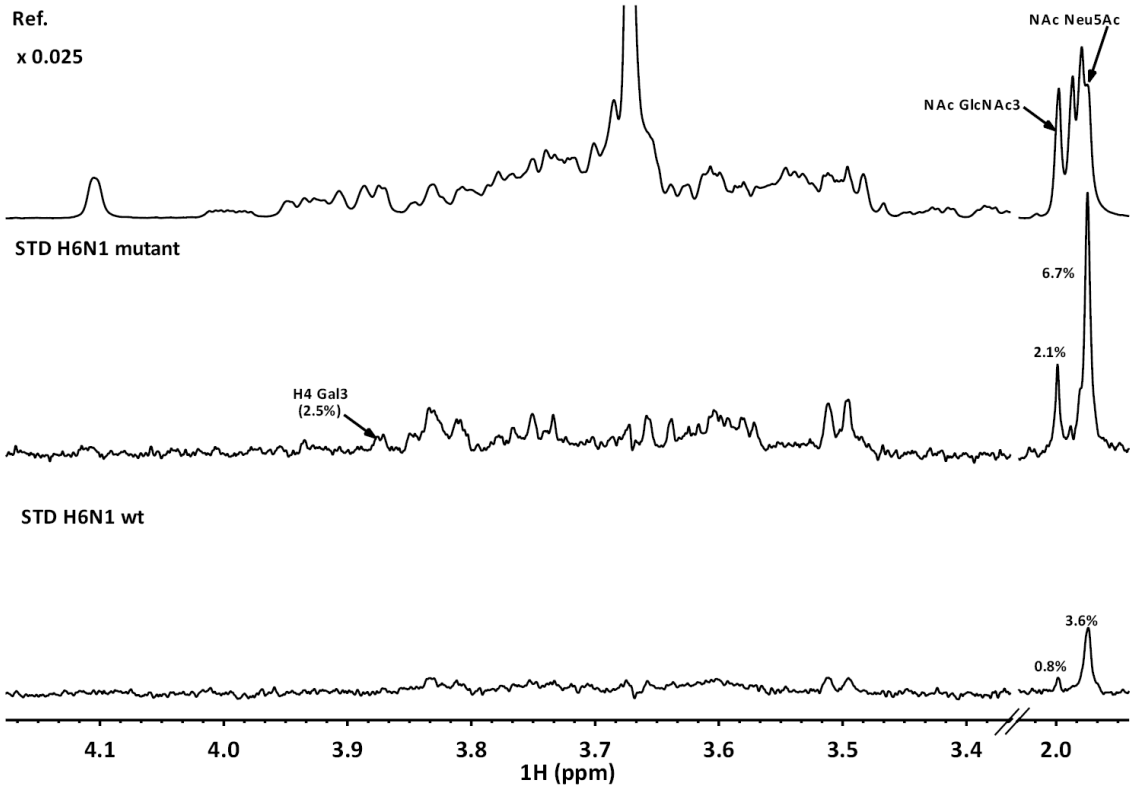


Figure 62. The interaction of glycan 3 with the two HA of the H6N1 strains. STD-NMR spectrum acquired at 600 MHz (298 K). The top panel shows the off-resonance spectrum. The middle panel contains the STD spectrum of 3 in the presence of the HA of the H6N1 mutant strain. The bottom panel displays the STD of glycan 3 in the presence of the H6N1 wild type. The saturation frequency was set up at δ 7 ppm (aromatic region) and the saturation time was 2 seconds. In both cases the same protein:ligand molar ratio (1:60) were used.

In particular, the STD-NMR spectrum of the H6N1 mutant strain showed a preferential binding towards the Neu5Ac end as described above for the H1N1 variant. The spectrum permits to discriminate the acetyl groups of the GlcNAc units (GlcNAc 1 and GlcNAc 2) from that the Neu5Ac. However, no signals for Neu5Ac H3a and H3b or for Gal 1 and Gal 2 H4 were observed. Only, H4 from Gal 3 was observed in the STD. Therefore, no evidence of the existence of an extended contact epitope along the full oligosaccharide chain was found in this case. For H6N1 WT, the STD signals were much less intense to those observed for the mutant human variant although a similar STD profile was observed for the acetyl region when compared to that of the mutant strain. Therefore, glycan **3** presents a much better affinity towards the HA of the mutant H6N1 strain.

Following similar methodology to that described above, glycan **3** was modeled in the binding site of the H6N1 HA mutant strain, using as starting point the available X-ray structure of H6N1 (PDB coordinates provided by personal communication with Prof. De Vries) (figure 63). The gt (ω dihedral angle) conformer of glycan **3** was docked to the crystal structure and the obtained complex was further minimized. The obtained 3D model fits with the obtained STD data,

where H4 from Gal 3 is fairly close to the surface of the protein, while H3a and H3b from Neu5Ac and H4 from Gal 2 are far away, at more than 4 Å of distance. In this case, the folded umbrella topology displayed by glycan **3** is slightly opened and places the reducing end of the oligosaccharide chain out of the protein binding pocket.

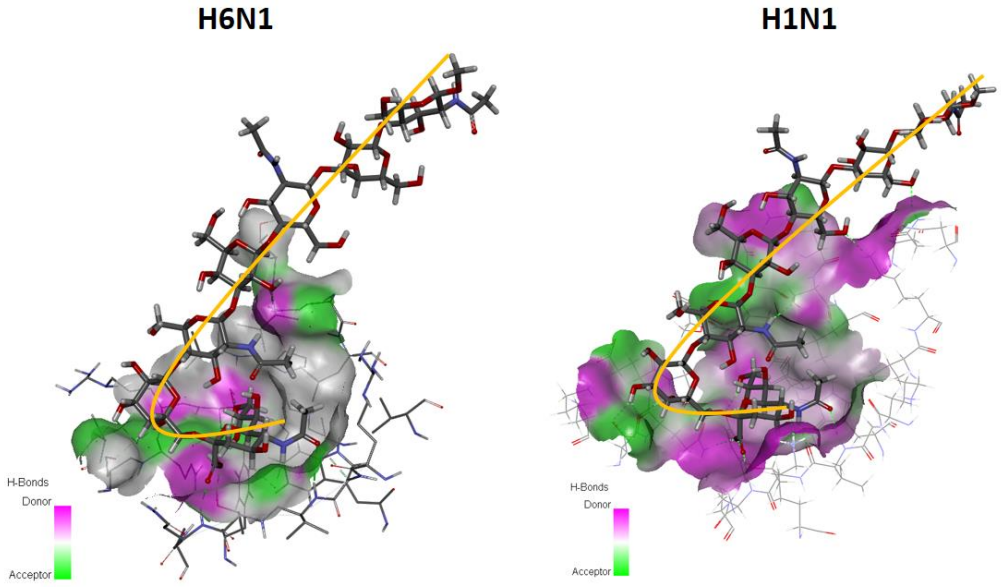
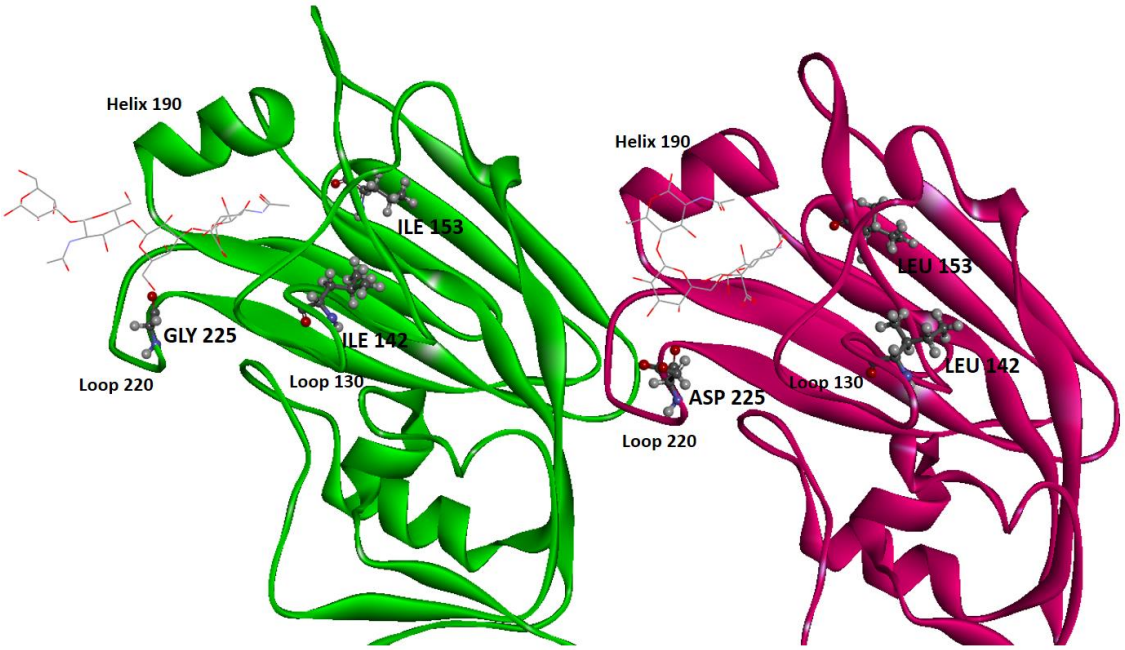


Figure 63. Proposed 3D model for the complex formed by the interaction of glycan 3 with the HA of the H6N1 mutant strain and comparison with the model for H1N1. The folded umbrella topology adopted by the glycan is shown, together with the interaction pattern.

Only three amino acids have changed in the H6N1 mutant with respect to the H6N1 WT strain: Leu153Ile, Leu142Ile and Gly225Asp (figure 64). Glycine 225 (WT) confers flexibility to the loop 220 and allows the α2-3 linkage oligosaccharide to be bound. However, its mutation into an Asp residue defines the presentation of loop 220, now hindering the binding of NeuNAcα2-3Gal moieties, but enabling the binding of the conformer of the Neu5Acα2-6Gal fragment, which forms the interacting epitope together with GlcNAc 3.



	153	142	225
H6N1 wt	LEU	LEU	GLY
H6N1 mutant	ILE	ILE	ASP

Figure 64. The mutated amino acids at the binding site of the HA of the H6N1 strain. In green, the wild type H6N1. In magenta, the H6N1 mutant strain.

An analogous study was performed for the H7N9 HA strains. The STD-NMR spectrum (figure 65) acquired for the complex of **3** with the mutant H7N9 strain presents a similar acetyl pattern to those observed for H6N1 and H1N1. However, the STD intensity of the GlcNAc 3 OAc group is marginal, significantly weaker than those observed before. For the wild type strain, the highest STD was much weaker than that measured for the mutant. Thus, the STD data also confirm that the wild type displays very weak affinity towards the α2-6 linkage, while the mutant strain is now able to recognize the Neu5Ac end of the glycan. However, there is not any extended epitope in this case and the experimental data strongly suggest that the reducing end of the glycan is pointing out of the protein.

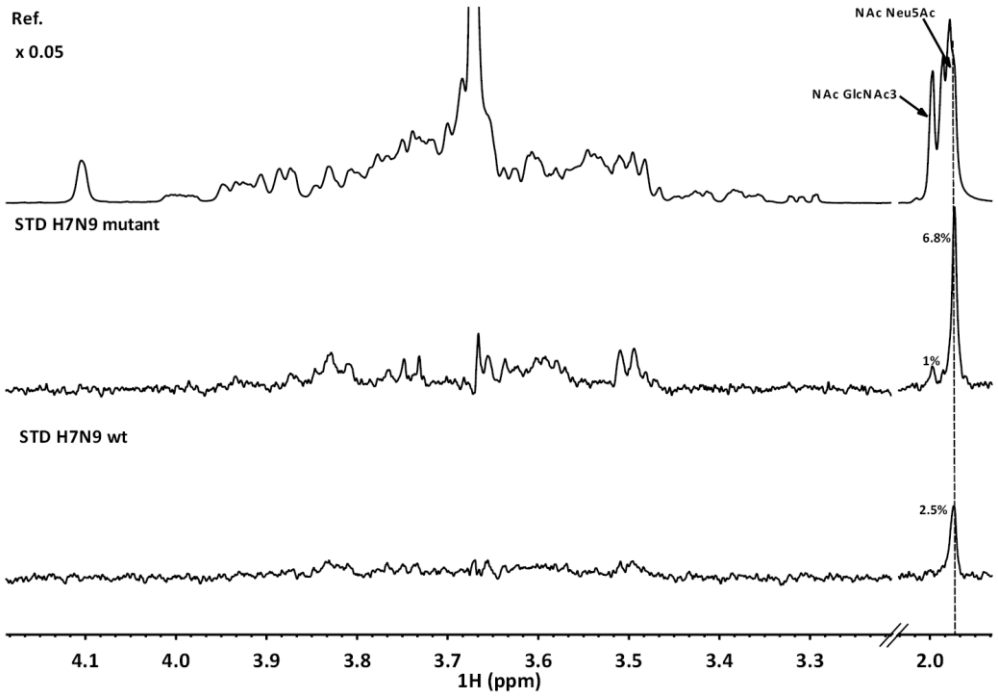


Figure 65. The interaction of glycan 3 with the two HA of the H7N9 strains. STD-NMR spectrum acquired at 600 MHz (298 K). The top panel shows the off-resonance spectrum. The middle panel contains the STD spectrum of 3 in the presence of the HA of the H7N9 mutant strain. The bottom panel displays the STD of glycan 3 in the presence of the H7N9 wild type. The saturation frequency was set up at δ 7 ppm (aromatic region) and the saturation time was 2 seconds.

A 3D model was built following the same methodology described above (figure 66), using the available crystal structure (PDB provided by Prof. De Vries) as a template for the gt conformer of the glycan.

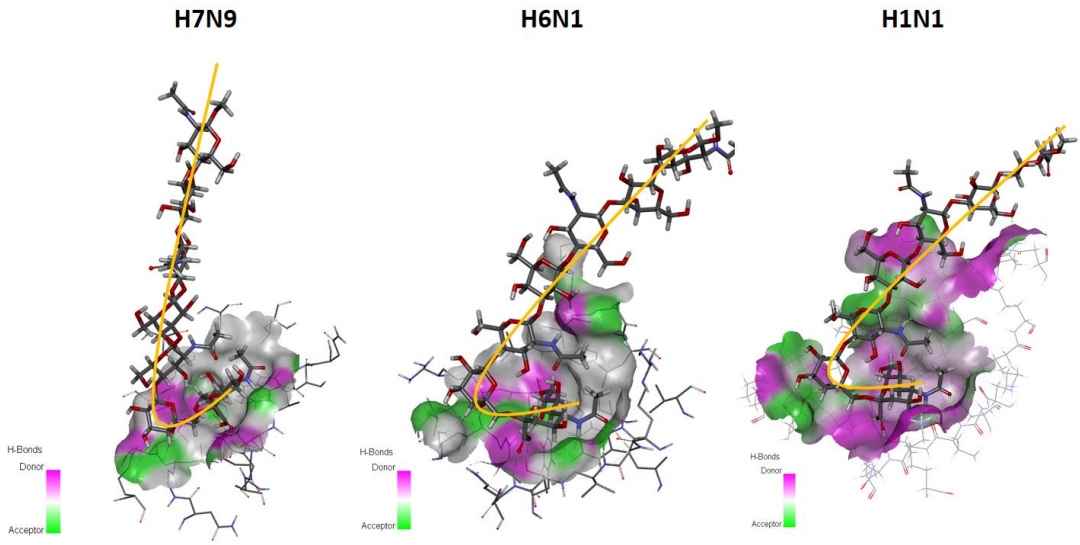
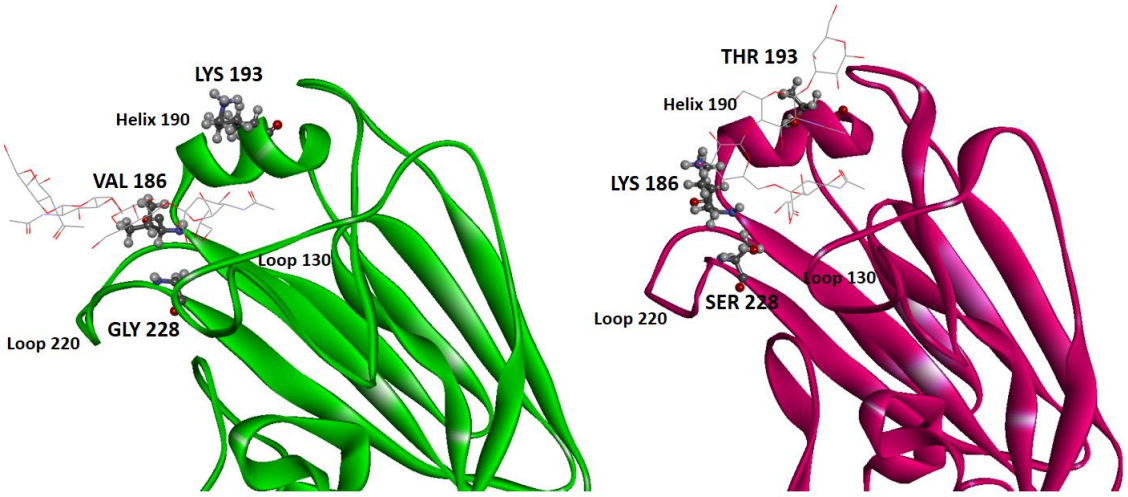


Figure 66. 3D model for the complex formed by the interaction of glycan 3 with the HA of the H7N9 mutant strain and comparison with the equivalent models with H6N1 and H1N1. A folded umbrella topology is adopted, although the glycan points out of the binding pocket. No extended epitope is present in this case.

Indeed, the generated 3D model showed that the glycan is only bound through the Neu5Ac non-reducing end, with the oligosaccharide chain pointing out of the pocket.

The mutations of only three amino acids (Val186Lys, Lys193Thr and Gly228Ser) at the binding site of HA dramatically switch the binding properties of the proteins (figure 67). Again, glycine 228 in the wild-type H7N9 provides the proper flexibility to the loop 220 to be able to accommodate the Neu5Acα2-3Gal-containing glycan. However, the Gly228Ser mutation is able to bind the α2-6 linked glycan. In fact, according to the 3D model, the new Ser228 and Lys186 amino acids establish intermolecular hydrogen bonds with the sialic acid moiety.



	186	193	228
H7N9 wt	VAL	LYS	GLY
H7N9 mutant	LYS	THR	SER

Figure 67. The mutated amino acids at the binding site of the H7N9 HA strain. In green, the wild type H7N9. In magenta, the H7N9 mutant strain.

3.5.2. Conclusions

In summary, the recognition features of glycans **2** and **3** when interacting with the HA of the H1N1 strain has been elucidated. Both molecules are recognized via the Neu5Ac moiety and present an extended epitope along the oligosaccharide chain, adopting a folded umbrella-like topology at the hemagglutinin binding site. Moreover, the mutation of three residues in the HAs of the H6N1 and H7N9 strains produce the switch of their binding properties from the selective selection of the Neu5Acα2-3Gal moiety to the recognition of the Neu5Acα2-6Gal analogue. However, the binding epitopes of the glycans rather different to those observed in

the H1N1 case and do not extend through the entire oligosaccharide chain, but only to the sialic acid residue and partially to its vicinal monosaccharides.

3.5.3. Experimental

Materials

All the HAs were provided by Dr. Robert De Vries form Utrecht University in the Netherlands. The H1N1 was the strain A/Kentucky/2007; H6N1 wild type and mutant, the A/Taiwan/1/2013 and the H7N9 wild type and mutant, the A/Shanghai2/2/2013.

Synthesis of the compounds

❖ General Methods

- **Gal extension**

The procedure described by Blixt *et al.*[123] was followed with minor modifications. Briefly, the initial mixtures contained the acceptor substrate (1 eq, 20-35 mM) and UDP-glucose (1.5 eq, 35-50 mM) in TRIS buffer (0.5-1 mL, 100 mM TRIS, 20 mM $MgCl_2$ pH 9). The reactions were initiated by addition of the fusion protein β4galT-galE[123] (15 mU based on the galactosyltransferase activity). The mixtures were slowly mixed at 37 °C in an incubator overnight. The pH was monitored and kept at a value of 9 by dropwise addition of NaOH. Once the reaction was observed to be completed by thin-layer chromatography (TLC) (see below), the reaction mixture was then ultrafiltered through a 10 kDa MWCO membrane to remove the proteins and passed through a Dowex cation resin to remove the excess of UDP-Glc and the released UDP and further purified by size exclusion chromatography (Sephadex P-2, 5×115 cm, H_2O). At all steps, the product was monitored by TLC performed on glass plates coated with Silica gel 60 (E. Merck) using the eluent system ethyl acetate:methanol:acetic acid:water (6:3:3:2 by volume) and visualized by treatment with 10% sulfuric acid in ethanol, followed by heating. The final products were checked by mass spectrometry data acquired with a MALDI-TOF (Applied Biosystems, DE) using 2,5-dihydroxy benzoic acid as matrix.

- **GlcNAc extension**

The initial mixtures contained the acceptor substrate (1 eq, 20-35 mM) and UDP-N-acetylglucosamine (1.5-2 eq, 30-60 mM) in sodium cacodylate buffer (0.5-1 mL, 100 mM Na cacodylate, 25 mM KCl, 1 mM DTT, 2 mM $MgCl_2$, pH 7.5). The reactions were initiated by addition of the enzyme β3GlcNAcT[124] (30 μL ≈ 1.5 mU) from *Helicobacter pylori*. The reaction mixtures were slowly mixed at 37 °C in an incubator for 96 h, and the pH was monitored and kept at a value of 7.5 by dropwise addition of NaOH. Once the reaction was completed, the reaction mixture was then ultrafiltered through a 10 kDa MWCO membrane to remove the proteins and passed through a Dowex cation resin to remove the excess of UDP-Glc and the released UDP and further purified by size exclusion chromatography (Sephadex G25, 5×115cm, H_2O). At all steps, the product was monitored by TLC, with two sequential elutions using the eluent system iso-propanol:water:ammonium hydroxide (6:3:2 by volume) on glass plates coated with Silica gel 60 (E. Merck) and visualized by treatment with 10% sulfuric acid in ethanol followed by heating. The final products were checked by mass spectrometry data acquired with a MALDI-TOF (Applied Biosystems, DE) using 2,5-dihydroxy benzoic acid as matrix

- **CMP-Neu5Ac production**

To a reaction mixture containing Neu5Ac (1 eq, 150 μmol, 50 μM) and CTP (1.2 eq, 180 μmol, 60 μM) in Tris buffer (100 mM TRIS, 20 mM $MgCl_2$, pH 9), CMP-Neu5Ac synthase[125] (50 μL) and commercial available inorganic pyrophosphatase (2.5 μL) were added. The mixtures were slowly mixed at 37 °C (310 K) in an incubator overnight, and the pH was monitored and kept at a value of 9 by dropwise addition of NaOH. The reaction was monitored by TLC performed on glass plates coated with Silica gel 60 (E. Merck) using the eluent system iso-propanol:water:ammonium hydroxide (6:3:2 by volume) and ethyl acetate:methanol:acetic acid:water (6:3:3:2) (the plate was ran once by each eluent system) and visualized by treatment with 10% sulfuric acid in ethanol followed by heating. No purification steps were necessary for further use of the product.

Enzyme expression

The expression of the 2,3-sialyltransferase PmST1 was adapted from the previously described protocol by Yu *et al.*[126] Briefly, the enzyme was expressed using Escherichia Coli BL21 (DE3),

purified in Ni^{2+}-NTA resin and directly used for the synthesis. The activity of the enzyme was determined by an enzymatic assay using radioactively tagged CMP-Neu5Ac.

Synthesis of compound **1**

The synthesis of the precursor of compound **1** (compound **1c**) was performed as described in the general methods section. Precursor **1c** (1 eq, 6.1 μmol, 30.5 mM) and CMP-Neu5Ac (1.5 eq, 9.15 μmol, 46 mM) were dissolved in 200 μL TRIS buffer (100 mM TRIS, 20 mM $MgCl_2$, pH 9). The sialyltransferase PmST1 previously expressed was then added (10 μL, 35 mU) to the reaction mixture. The reaction was mixed at 37 °C in an incubator for 4 days, and the pH was monitored and kept at 9 by dropwise addition of NaOH. The product was monitored by TLC performed on glass plates coated with Silica gel 60 (E. Merck) using the eluent system ethyl acetate:methanol:acetic acid:water (3:3:3:2 by volume) and visualized by treatment with 10% sulfuric acid in ethanol followed by heating. The reaction mixture was ultrafiltered through a 10 kDa MWCO membrane to remove the sialyltransferase and passed through a Dowex cation resin to remove the excess of CMP-Neu5Ac and the released CMP and further purified by size exclusion chromatography (Sephadex G25, 5×130 cm, H_2O). The final product in 62 % yield in the last reaction step was checked by NMR (see assignment section) and MS (MALDI-TOF) m/z $[M + H]^+$ = 1131 and $[M + Na]^+$ = 1153 (figure 68).

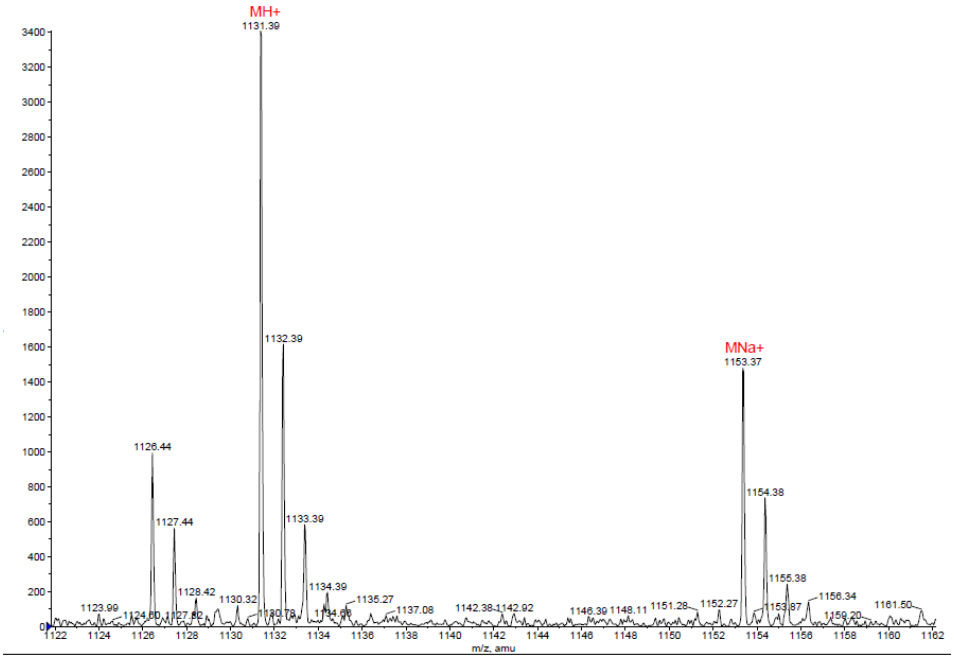


Figure 68. MALDI-TOF spectrum of compound 1.

Synthesis of compound **2**

Precursor **1c** (1 eq, 6.1 μmol, 50 mM) and CMP-Neu5Ac (2 eq, 12 μmol, 100 mM) were dissolved in 120 μL of sodium cacodylate buffer (100 mM sodium cacodylate, 20 mM $MgCl_2$, pH 8.5). The sialyltransferase hST6Gal[127] was then added (10 μL, 11.2 mU) to the reaction mixture. The reaction was mixed at 37 °C (310 K) in an incubator 7 days, and the pH was monitored and kept at 8.5 by dropwise addition of NaOH. The product was monitored by TLC performed on glass plates coated with Silica gel 60 (E. Merck) using the eluent system ethyl acetate:methanol:acetic acid:water (3:3:3:2 by volume) and visualized by treatment with 10% sulfuric acid in ethanol followed by heating. The reaction mixture was ultrafiltered through a 10 kDa MWCO membrane to remove the sialyltransferase and passed through a Dowex cation resin to remove excess of CMP-Neu5Ac and the released CMP and further purified by size exclusion chromatography (Biogel P-2, 5×115 cm, H_2O). The final product was checked by NMR (see assignment section) and MS (MALDI-TOF) m/z $[M + H]^+$ = 1132 and $[M + Na]^+$ = 1154 (figure 69).

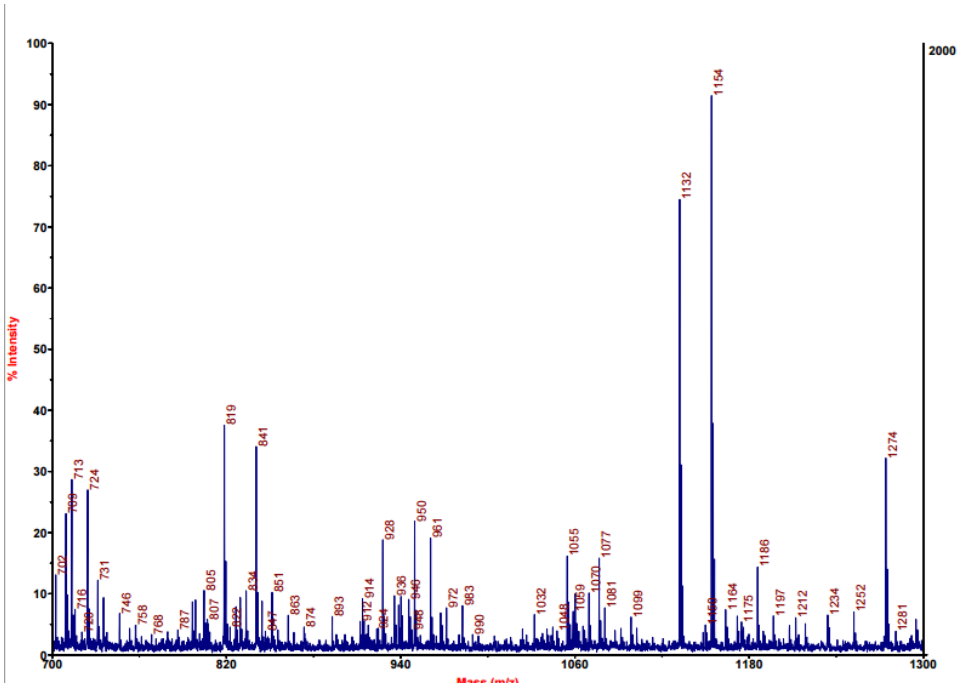


Figure 69. MALDI-TOF spectrum of compound 2.

Synthesis of compound **3**

Precursor **3b** (1 eq, 5.9 μmol, 50 mM) and CMP-Neu5Ac (2 eq, 12 μmol, 100 mM) were dissolved in 120 μL of sodium cacodylate buffer (100 mM sodium cacodylate, 20 mM $MgCl_2$,

pH 8.5). The sialyltransferase was then added (10 μL, 11.2 mU) to the reaction mixture. The reaction was mixed at 37 °C (310 K) in an incubator 5 days, and the pH was monitored and kept at 8.5 by dropwise addition of NaOH. The product was monitored by TLC performed on glass plates coated with Silica gel 60 (E. Merck) and visualized by treatment with 10% sulfuric acid in ethanol followed by heating using the eluent system ethyl acetate:methanol:acetic acid:water (3:3:3:2 by volume). The reaction mixture was ultrafiltered through a 10 kDa MWCO membrane to remove the sialyltransferase and passed through a Dowex cation resin to remove the excess of CMP-Neu5Ac and the released CMP and further purified by size exclusion chromatography (Sephadex G25, 5×115 cm, H_2O). The final product was checked by NMR (see assignment section) and MS (MALDI-TOF) m/z $[M + H]^+$ = 1497 and $[M + Na]^+$ = 1519 (figure 70).

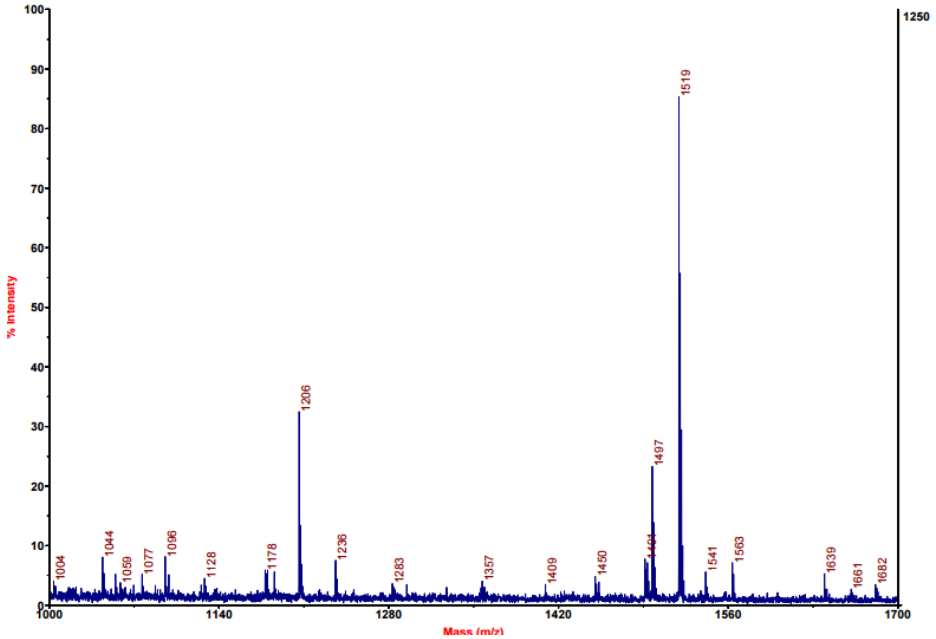


Figure 70. MALDI-TOF spectrum of compound 3.

Synthesis of compound **4**

Compound **4a** (1 eq, 8.3 μmol, 20 mM) was dissolved in 410 μL of a 62.5 mM stock solution of sodium ethylacetate. The comercially available AUS Neuraminidase (from Sigma Aldrich) (10 μL) was added to the reaction mixture and then mixed in an incubator at 37ºC (310 K) for 2 days. The reaction was worked it up by ultrafiltration through a 10 kDa MWCO membrane to remove the sialyltransferase and passed through a Dowex cation resin to remove excess of CMP-Neu5Ac and the released CMP and further purified by size exclusion chromatography

(Sephadex G25, 5×115cm, H_2O). The product (compound **4b**) was monitored by TLC performed on glass plates coated with Silica gel 60 (E. Merck) and visualized by treatment with 10% sulfuric acid in ethanol followed by heating using the eluent system isopropanol: water: ammonium hydroxide (6:3:2 by volume). The extension of the LacNAc units was performed as described in the general methods sections to finally obtain compound **4d.** The precursor **4d** (1 eq, 2 μmol, 40 mM) and CMP-Neu5Ac (4 eq, 8 μmol, 160 mM) were dissolved in 50 μL of sodium cacodylate buffer (100 mM sodium cacodylate, 20 mM $MgCl_2$, pH 8.5). The sialyltransferase was then added (7 μL) to the reaction mixture. The reaction was mixed at 37 °C (310 K) in an incubator 7 days, and the pH was monitored and kept at 8.5 by dropwise addition of NaOH. The product was monitored by TLC performed on glass plates coated with Silica gel 60 (E. Merck) and visualized by treatment with 10% sulfuric acid in ethanol followed by heating using the eluent system isopropanol: water: ammonium hydroxide (6:3:2 by volume). The reaction mixture was ultrafiltered through a 10 kDa MWCO membrane to remove sialyltransferase and passed through a Dowex cation resin to remove the excess of CMP-Neu5Ac and the released CMP and further purified by size exclusion chromatography (Sephadex G25, 5×115 cm, H_2O). The final product was checked by MS (MALDI-TOF) m/z $[M + H]^+$ = 3136 and $[M + Na]^+$ = 3158 (figure 71).

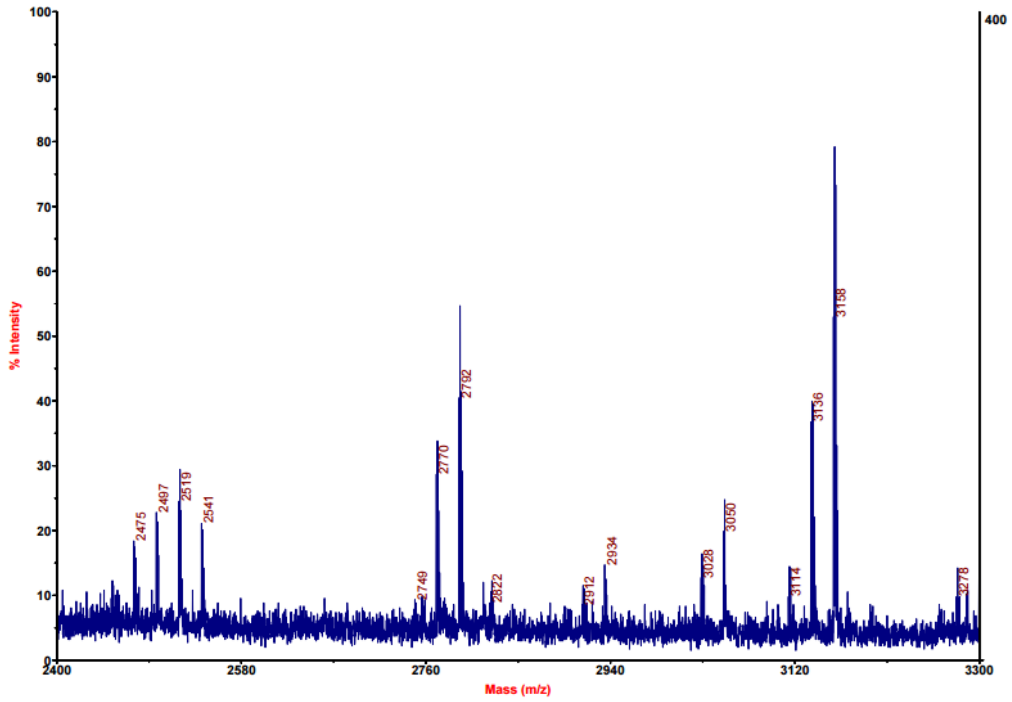


Figure 71. MALDI-TOF spectrum of compound 4a.

Synthesis of compound **5**

From compound **4b**, described above, the extension of the LacNAc units to finally obtain compound **5b** was performed as described in the general methods section and checked by MS (MALDI-TOF) m/z $[M + H]^+$ = 3240 and $[M + Na]^+$ = 3262 (figure 72). Then precursor **5b** (1 eq, 1.7 μmol, 34 mM) and CMP-Neu5Ac (4 eq, 6.8 μmol, 136 mM) were dissolved in 50 μL of sodium cacodylate buffer (100 mM sodium cacodylate, 20 mM $MgCl_2$, pH 8.5). The sialyltransferase was then added (7 μL) to the reaction mixture. The reaction was mixed at 37 °C (310 K) in an incubator 7 days, and the pH was monitored and kept at 8.5 by dropwise addition of NaOH. The product was monitored by TLC performed on glass plates coated with Silica gel 60 (E. Merck) and visualized by treatment with 10% sulfuric acid in ethanol followed by heating using the eluent system isopropanol: water: ammonium hydroxide (6:3:2 by volume). The reaction mixture was ultrafiltered through a 10 kDa MWCO membrane to remove sialyltransferase and passed through a Dowex cation resin to remove excess of CMP-Neu5Ac and the CMP released and further purified by size exclusion chromatography (Sephadex G25, 5×115 cm, H_2O). The purification of compound **5** could not be finished during the internship and therefore, the MS spectrum could not be obtained.

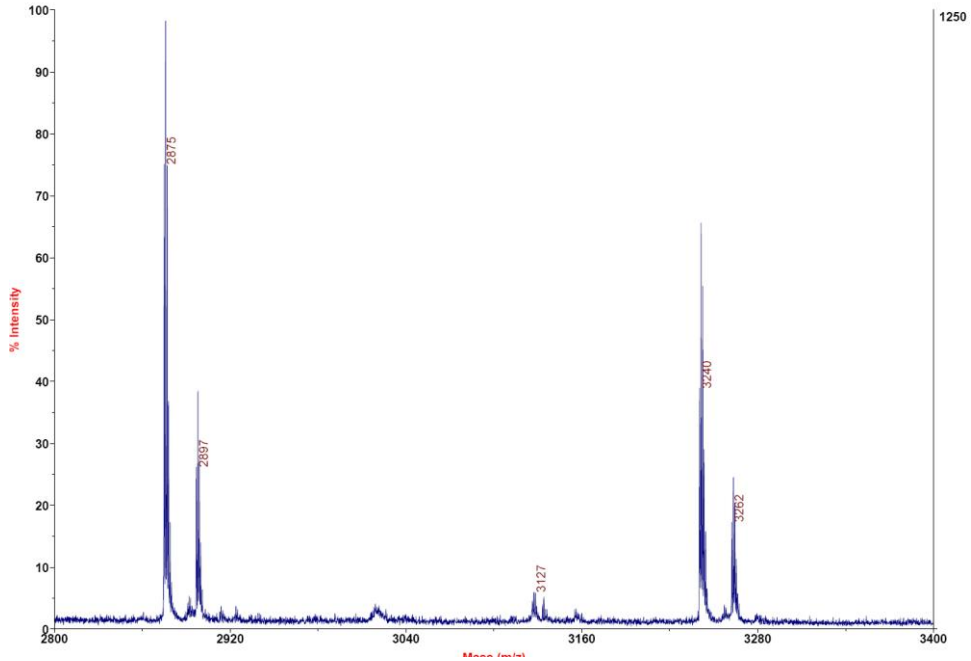
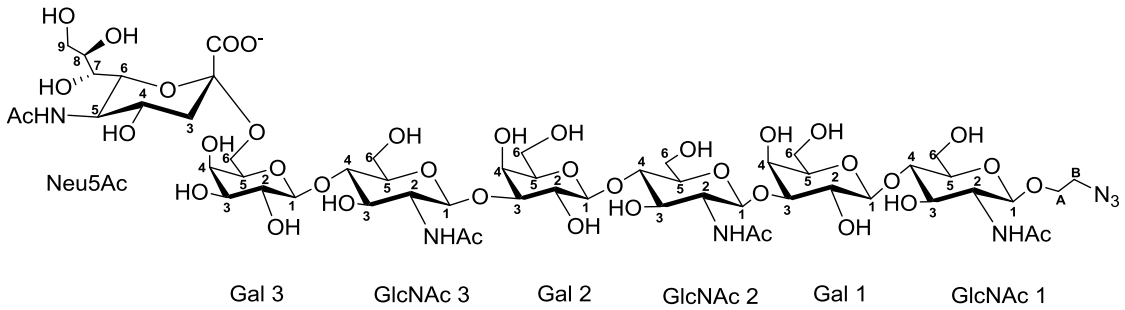


Figure 72. MALDI-TOF spectrum of compound 5b.

NMR experiments

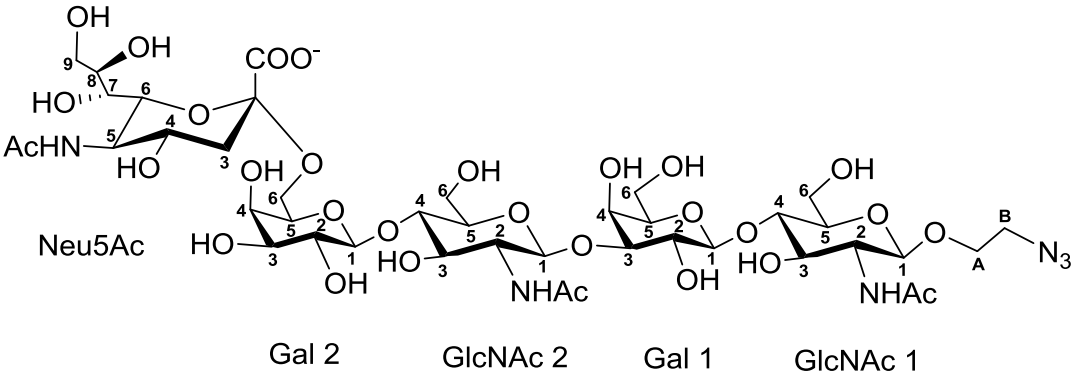
The ^{1}H NMR resonances of glycans **1**, **2** and **3** were assigned by the acquisition of a set of standard homo- and hetero-nuclear NMR experiments: 1D ^{1}H-NMR, 2D TOCSY at different

mixing times and 2D heteronuclear ^{1}H-^{13}C HSQC and ^{1}H-^{13}C HSQC–TOCSY in D_2O in a 600 MHz Bruker Avance spectrometer, equipped with a cryoprobe at 298 K. The assignments are given below.



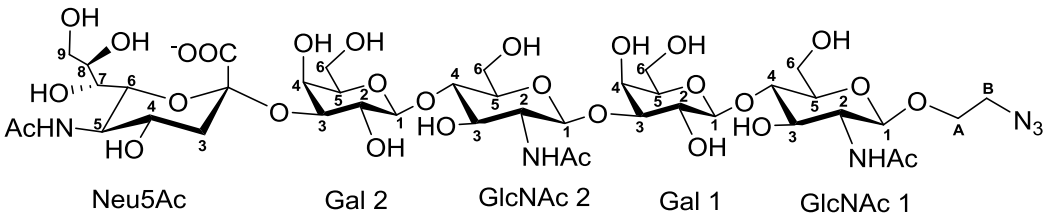
$^{1}H/^{13}C$ (δ, ppm)	Neu5Ac	Gal 3	GlcNAc 3	Gal 2
1		4.37/103.4	4.38/102.5	4.38/102.8
2		3.46/70.9	3.72/55.2	3.51/70.0
3	1.64, 2.59/40.1	3.62/72.6	3.65/72.4	3.64/72.3
4	3.58/68.4	3.85/68.6	3.64/78.1	4.08/68.5
5	3.72/52.1	3.74/73.7	3.52/74.5	3.64/74.9
6	3.62/72.6	3.47, 3.91/63.3	3.76, 3.88/60.0	3.65, 3.69/61.1
7	3.48/68.5			
8	3.81/72.0			
9	3.57, 3.80/62.7			
NH	7.84		8.08	
NAc	1.94/22.1		1.97/22.40	

$^{1}H/^{13}C$ (δ , ppm)	GlcNAc 2	Gal 1	GlcNAc 1	A
1	4.35/102.5	4.38/102.8	4.52/100.9	3.97, 3.69/68.7
2	3.72/55.2	3.51/70.0	3.71/55.1	
3	3.65/72.4	3.64/72.3	3.72/72.3	**B**
4	3.63/78.4	4.08/68.5	3.59/80.4	3.36, 3.40/50.3
5	3.50/74.6	3.64/74.9	3.52/74.8	
6	3.76, 3.88/60.0	3.65, 3.69/61.1	3.76, 3.88/60.0	
NH	8.10		8.12	
NAc	1.95/22.3		1.96/22.3	



$^{1}H/^{13}C$ (δ , ppm)	Neu5Ac	Gal 2	GlcNAc 2
1		4.47	4.37
2		3.46/71.0	3.72/54.9
3	2.60, 1.65/40.3	3.61/72.5	3.71/72.4
4	3.66/69.5	3.86/69.4	3.61/78.6
5	3.73/52.1	3.74/73.6	3.52/74.6
6	3.65/72.4	3.47, 3.91/63.4	3.76, 3.89/60.3
7	3.52/69.7		
8	3.81/71.8		
9	3.57, 3.80/62.9		
NAc	1.96/22.5		1.98/22.5

$^{1}H/^{13}C$ (δ , ppm)	Gal 1	GlcNAc 1	A
1	4.51	4.61	3.63;3.81/68.2
2	3.51/70.2	3.67/54.9	
3	3.66/82.0	3.70/72.4	**B**
4	4.08/69.5	3.61/78.6	3.50, 3.44/47.5
5	3.64/74.9	3.52/74.6	
6	3.66, 3.69/61.0	3.76, 3.86/60.3	
NAc		1.98/22.5	



$^{1}H/^{13}C$ (δ , ppm)	Neu5Ac	Gal 2	GlcNAc 2
1		4.48/102.4	-
2		3.53/68.2	3.70/54.9
3	1.73, 2.69/39.4	4.03/75.5	3.69/72.0
4	3.62/68.6	3.86/67.6	3.64/78.4
5	3.77/51.6	3.64/75.2	3.51/74.5
6	3.56/72.9	3.70, 3.65/61.0	3.79, 3.89/59.9
7	3.53/68.2		
8	3.82/71.9		
9	3.79, 3.57/62.6		
NH	7.98		8.2
NAc	1.96/22.4		1.96/22.4

$^{1}H/^{13}C$ (δ , ppm)	Gal 1	GlcNAc 1	A
1	4.39/103.0	4.53/101.1	3.70, 3.97/68.7
2	3.52/69.7	3.73/55.3	
3	3.65/82.1	3.66/72.2	**B**
4	4.08/68.4	3.67/77.9	3.41, 3.36/50.4
5	3.67/75.2	3.53/74.7	
6	-	3.92, 3.75/60.2	
NH		8.18	
NAc		1.97/22.4	

STD-NMR experiments were performed on the complexes of the different HAs and the different glycans using PBS (pH 7.3) deuterated buffered solutions. For the case of H1N1 HA, a sample of 3.5 μM of HA and 300 μM of glycan **2** (molar ratio 1:86) was prepared. The same molar ratio HA:ligand (1:86) was used for the case of H1N1 HA (3.5 μM) and glycan **3** (300 μM). For the H6N1 (wild type and mutant strains) and H7N9 strains (wild type and mutant), a solution of 5 μM of HA and 300 μM of glycan **3** was employed (molar ratio 1:60). The STD experiments were acquired at 298 K on a Bruker Avance 600 MHz spectrometer equipped with

a cryoprobe; the saturation frequency was set up at δ 7 ppm (aromatic region) during 2 seconds.

Computational model

The initial conformation of glycan **3** was obtained using the GLYCAM online carbohydrate builder.[119,128] Of all possible structures, the most stable gauche-trans (gt) conformation around the ω angle of the terminal Gal residue, to which the sialic acid moiety is linked, was taken as starting geometry. The available X-ray crystal structures of H1N1 and the wild types of H6N1 and H7N9 were taken from the Protein Data Bank[70] (H1N1 ID: 4JTV; H6N1 wt ID: 4YYA; H7N9 wt ID: 4N63). The coordinates of the H6N1 and H7N9 mutants were provided by Dr. De Vries from Utrecht University in the Netherlands. Then, the corresponding glycan-HA complexes were built and minimized (5000 iterations) using the Maestro program (Schrodinger Computational Suite) and employing the AMBER* force field in implicit water solvent. The macromolecule was considered to be fixed during the minimization process, while the ligand was left flexible.

3.6. References

(1) Mammen, M.; Choi, S. K.; Whitesides, G. M. *Angew. Chem. Int. Ed.* **1998**, *37* (20), 2754.
(2) Fasting, C.; Schalley, C. A.; Weber, M.; Seitz, O.; Hecht, S.; Koksch, B.; Dernedde, J.; Graf, C.; Knapp, E. W.; Haag, R. *Angew. Chem. Int. Ed.* **2012**, *51* (42), 10472.
(3) Ribeiro-Viana, R.; Sánchez-Navarro, M.; Luczkowiak, J.; Koeppe, J. R.; Delgado, R.; Rojo, J.; Davis, B. G. *Nat. Commun.* **2012**, *3*, 1303.
(4) R. Moreau, J. Dausset, J. Bernard, J. M. *Bull. Mem. Soc. Med. Hop. Paris* **1957**, *73*, 569 – 587.
(5) Springer, G. F.; Desai, P. R.; Banatwala, I. *Naturwissenschaften* **1974**, *61* (10), 457–458.
(6) Cartron, J.P.; Andreu, G.; Cartron, J.; Bird, G. W. G.; Salmon, C.; Gerbal, A. *Eur. J. Biochem.* **1978**, *92* (1), 111.
(7) Cartron, J. P.; Andreu, G.; Cartron, J.; Salmon, C. H.; Bird, G. W. G. *Lancet* **2017**, *311* (8069), 856–857.
(8) Thurnher, M.; Berger, E. G.; Clausen, H.; Fierz, W.; Lanzavecchia, A. *Eur. J. Immunol.* **1992**, *22* (7), 1835.
(9) Springer, G. F. *J. Mol. Med.* **1997**, *75* (8), 594.
(10) Springer, G. F. *Science (80-.).* **1984**, *224* (4654), 1198.
(11) Desai, P. R. *Transfus. Med. Rev.* **2017**, *14* (4), 312.
(12) Kakeji, Y.; Tsujitani, S.; Mori, M.; Maehara, Y.; Sugimachi, K. *Cancer* **1991**, *68* (11), 2438.
(13) Hansen, J. E.; Nielsen, C.; Arendrup, M.; Olofsson, S.; Mathiesen, L.; Nielsen, J. O.; Clausen, H. *J. Virol.* **1991**, *65* (12), 6461.

(14) Xie, L.; Tan, C.; Fan, J.; Fu, P.; Tang, Y.; Tao, Y.; Qin, W. *Int. Urol. Nephrol.* **2013**, *45* (2), 571.
(15) Beatson, R.; Maurstad, G.; Picco, G.; Arulappu, A.; Coleman, J.; Wandell, H. H.; Clausen, H.; Mandel, U.; Taylor-Papadimitriou, J.; Sletmoen, M.; Burchell, J. M. *PLoS One* **2015**, *10* (5), 1.
(16) Fu, C.; Zhao, H.; Wang, Y.; Cai, H.; Xiao, Y.; Zeng, Y.; Chen, H. *HLA*. **2016**, 88, 275.
(17) Chia, J.; Goh, G.; Bard, F. *Biochim. Biophys. Acta* **2016**, *1860* (8), 1623.
(18) Saeland, E.; Van Vliet, S. J.; Bäckström, M.; Van Den Berg, V. C. M.; Geijtenbeek, T. B. H.; Meijer, G. A.; Van Kooyk, Y. *Cancer Immunol. Immunother.* **2007**, *56* (8), 1225.
(19) Ju, T.; Wang, Y.; Aryal, R. P.; Lehoux, S. D.; Ding, X.; Kudelka, M. R.; Cutler, C.; Zeng, J.; Wang, J.; Sun, X.; Heimburg-Molinaro, J.; Smith, D. F.; Cummings, R. D. *Proteomics Clin. Appl.* **2013**, *7* (9–10), 618.
(20) Napoletano, C.; Rughetti, A.; Agervig Tarp, M. P.; Coleman, J.; Bennett, E. P.; Picco, G.; Sale, P.; Denda-Nagai, K.; Irimura, T.; Mandel, U.; Clausen, H.; Frati, L.; Taylor-Papadimitriou, J.; Burchell, J.; Nuti, M. *Cancer Res.* **2007**, *67* (17), 8358.
(21) Galan, M. C.; Dumy, P.; Renaudet, O. *Chem. Soc. Rev.* **2013**, *42* (11), 4599.
(22) Richichi, B.; Thomas, B.; Fiore, M.; Bosco, R.; Qureshi, H.; Nativi, C.; Renaudet, O.; BenMohamed, L. *Angew. Chem. Int. Ed.* **2014**, *53* (44), 11917.
(23) Paulsen, H.; Brockhausen, I. *Glycoconj. J.* **2001**, *18* (11–12), 867.
(24) Ohyama, C. *Int. J. Clin. Oncol.* **2008**, *13* (4), 308.
(25) Ardá, A.; Bosco, R.; Sastre, J.; Cañada, F. J.; André, S.; Gabius, H.-J.; Richichi, B.; Jiménez-Barbero, J.; Nativi, C. *Eur. J. Org. Chem.* **2015**, *2015* (31), 6823.
(26) Coltart, D. M.; Royyuru, A. K.; Williams, L. J.; Glunz, P. W.; Sames, D.; Kuduk, S. D.; Schwarz, J. B.; Chen, X.-T.; Danishefsky, S. J.; Live, D. H. *J. Am. Chem. Soc.* **2002**, *124* (33), 9833.
(27) Tuccillo, F. M.; De Laurentiis, A.; Palmieri, C.; Fiume, G.; Bonelli, P.; Borrelli, A.; Tassone, P.; Scala, I.; Buonaguro, F. M.; Quinto, I.; Scala, G. *Biomed Res. Int.* **2014**, *2014, 1*.
(28) Wu, A. M.; Wu, J. H.; Tsai, M. S.; Yang, Z.; Sharon, N.; Herp, A. *Glycoconj. J.* **2007**, *24* (9), 591.
(29) Rambaruth, N. D.; Greenwell, P.; Dwek, M. V. *Glycobiology* **2012**, *22* (6), 839.
(30) Turton, K.; Natesh, R.; Thiyagarajan, N.; Chaddock, J. A.; Acharya, K. R. *Glycobiology* **2004**, *14* (10), 923.
(31) Iglesias, J. L.; Lis, H.; Sharon, N. *Eur. J. Biochem.* **1982**, *123* (2), 247.
(32) Lis, H.; Joubert, F. J.; Sharon, N. *Phytochemistry* **1985**, *24* (12), 2803.
(33) Gabius, H.-J.; Kaltner, H.; Kopitz, J.; André, S. *Trends Biochem. Sci.* **2017**, *40* (7), 360.
(34) Zelensky, A. N.; Gready, J. E. *FEBS J.* **2005**, *272* (24), 6179.
(35) Marcelo, F.; Garcia-Martin, F.; Matsushita, T.; Sardinha, J.; Coelho, H.; Oude-Vrielink, A.; Koller, C.; André, S.; Cabrita, E. J.; Gabius, H. J.; Nishimura, S. I.; Jiménez-Barbero, J.; Cañada, F. J. *Chem. - A Eur. J.* **2014**, *20* (49), 16147.
(36) Mortezai, N.; Behnken, H. N.; Kurze, A. K.; Ludewig, P.; Buck, F.; Meyer, B.; Wagener, C. *Glycobiology* **2013**, *23* (7), 844.
(37) Prokop, O.; Uhlenbruck, G. *Med. Welt* **1969**, *46*, 2515.
(38) Ju, T.; Otto, V. I.; Cummings, R. D. *Angew. Chem. Int. Ed.* **2011**, *50* (8), 1770.
(39) Ardá, A.; Bosco, R.; Sastre, J.; Cañada, F. J.; André, S.; Gabius, H. J.; Richichi, B.; Jiménez-Barbero, J.; Nativi, C. *Eur. J. Org. Chem.* **2015**, *2015* (31), 6823.

(40) Langer, H. F.; Chavakis, T. *J. Cell. Mol. Med.* **2009**, *13* (7), 1211.
(41) Unione, L.; Gimeno, A.; Valverde, P.; Calloni, I.; Coelho, H.; Mirabella, S.; Poveda, A.; Jiménez-Barbero, A. A. and J. *Current Medicinal Chemistry*. **2017**, 1.
(42) Barondes, S. H.; Castronovo, V.; Cooper, D. N. W.; Cummings, R. D.; Drickamer, K.; Felzi, T.; Gitt, M. A.; Hirabayashi, J.; Hughes, C.; Kasai, K. ichi; Leffler, H.; Liu, F. T.; Lotan, R.; Mercurio, A. M.; Monsigny, M.; Pillai, S.; Poirer, F.; Raz, A.; Rigby, P. W. J.; Rini, J. M.; Wang, J. L. *Cell* **1994**, *76* (4), 597.
(43) Cooper, D. N. W. *Biochim. Biophys. Acta - Gen. Subj.* **2002**, *1572* (2–3), 209.
(44) Leffler, H.; Carlsson, S.; Hedlund, M.; Qian, Y.; Poirier, F. *Glycoconj. J.* **2002**, *19* (7), 433.
(45) Hoyer, K. K.; Pang, M.; Gui, D.; Shintaku, I. P.; Kuwabara, I.; Liu, F.-T.; Said, J. W.; Baum, L. G.; Teitell, M. A. *Am. J. Pathol.* **2017**, *164* (3), 893.
(46) Takenaka, Y.; Fukumori, T.; Yoshii, T.; Oka, N.; Inohara, H.; Kim, H.-R. C.; Bresalier, R. S.; Raz, A. *Mol. Cell. Biol.* **2004**, *24* (10), 4395.
(47) Choi, J.-H.; Chun, K.-H.; Raz, A.; Hong, W. K.; Lotan, R. *Cancer Biol. Ther.* **2004**, *3* (5), 447.
(48) Yang, R. Y.; Hsu, D. K.; Liu, F. T. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (13), 6737.
(49) Wang, J. L.; Gray, R. M.; Haudek, K. C.; Patterson, R. J. *Biochim. Biophys. Acta - Gen. Subj.* **2004**, *1673* (1–2), 75.
(50) Califice, S.; Castronovo, V.; Bracke, M.; van den Brule, F. *Oncogene* **2004**, *23* (45), 7527.
(51) Fukumori, T.; Takenaka, Y.; Oka, N.; Yoshii, T.; Hogan, V.; Inohara, H.; Kanayama, H.; Kim, H.-R. C.; Raz, A. *Cancer Res.* **2004**, *64* (10), 3376.
(52) Nangia-Makker, P.; Nakahara, S.; Hogan, V.; Raz, A. *J. Bioenerg. Biomembr.* **2007**, *39* (1), 79.
(53) Ahmad, N.; Gabius, H.-J.; André, S.; Kaltner, H.; Sabesan, S.; Roy, R.; Liu, B.; Macaluso, F.; Brewer, C. F. *J. Biol. Chem.* **2004**, *279* (12), 10841.
(54) Gabius, H.-J.; André, S.; Jiménez-Barbero, J.; Romero, A.; Solís, D. *Trends Biochem. Sci.* **2017**, *36* (6), 298.
(55) Herrmann, J.; Turck, C. W.; Atchison, R. E.; Huflejt, M. E.; Poulter, L.; Gitt, M. A.; Burlingame, A. L.; Barondes, S. H.; Leffler, H. *J. Biol. Chem.* **1993**, *268* (35), 26704.
(56) Butler, G. S.; Dean, R. A.; Tam, E. M.; Overall, C. M. *Mol. Cell. Biol.* **2008**, *28* (15), 4896.
(57) Puthenedam, M.; Wu, F.; Shetye, A.; Michaels, A.; Rhee, K.-J.; Kwon, J. H. *Inflamm. Bowel Dis.* **2011**, *17* (1), 260.
(58) Kopitz, J.; von Reitzenstein, C.; André, S.; Kaltner, H.; Uhl, J.; Ehemann, V.; Cantz, M.; Gabius, H.-J. *J. Biol. Chem.* **2001**, *276* (38), 35917.
(59) Ochieng, J.; Fridman, R.; Nangia-Makker, P.; Kleiner, D. E.; Liotta, L. a; Stetler-Stevenson, W. G.; Raz, a. *Biochemistry* **1994**, *33* (47), 14109.
(60) C. Nativi, B. Richichi, S. R. *RSC Drug Discov. Ser. No. 43 - Carbohydrates Drug Des. Discov.* **2015**, *10*, 242.
(61) Lovejoy, B.; Welch, A. R.; Carr, S.; Luong, C.; Broka, C.; Hendricks, R. T.; Campbell, J. A.; Walker, K. A. M.; Martin, R.; Van Wart, H.; Browner, M. F. *Nat. Struct. Biol.* **1999**, *6* (3), 217.
(62) Hajduk, P. J.; Sheppard, G.; Nettesheim, D. G.; Olejniczak, E. T.; Shuker, S. B.; Meadows, R. P.; Steinman, D. H.; Carrera, G. M.; Marcotte, P. A.; Severin, J.; Walter, K.; Smith, H.; Gubbins, E.; Simmer, R.; Holzman, T. F.; Morgan, D. W.; Davidsen, S. K.; Summers, J. B.; Fesik, S. W. *J. Am. Chem. Soc.* **1997**, *119* (25), 5818.
(63) MacPherson, L. J.; Bayburt, E. K.; Capparelli, M. P.; Carroll, B. J.; Goldstein, R.; Justice,

M. R.; Zhu, L.; Hu, S.; Melton, R. A.; Fryer, L.; Goldberg, R. L.; Doughty, J. R.; Spirito, S.; Blancuzzi, V.; Wilson, D.; O'Byrne, E. M.; Ganu, V.; Parker, D. T. *J. Med. Chem.* **1997**, *40* (16), 2525.

(64) Bartoloni, M.; Domínguez, B. E.; Dragoni, E.; Richichi, B.; Fragai, M.; André, S.; Gabius, H. J.; Ardá, A.; Luchinat, C.; Jiménez-Barbero, J.; Nativi, C. *Chem. - A Eur. J.* **2013**, *19* (6), 1896.

(65) Dalvit, C.; Piotto, M. *Magn. Reson. Chem.* **2017**, *55* (2), 106.

(66) Dalvit, C.; Flocco, M.; Veronesi, M.; Stockman, B. J. *Combinatorial Chemistry & High Throughput Screening*. **2002**, 605.

(67) P. McIntosh, L. In *Encyclopedia of Biophysics*; Roberts, G. C. K., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, **2013**; 386.

(68) Groves, P.; Rasmussen, M. O.; Molero, M. D.; Samain, E.; Cañada, F. J.; Driguez, H.; Jiménez-Barbero, J. *Glycobiology* **2004**, *14* (5), 451.

(69) Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Kent Wenger, R.; Yao, H.; Markley, J. L. *Nucleic Acids Res.* **2008**, *36* (suppl_1), D402.

(70) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28* (1), 235.

(71) Bertini, I.; Calderone, V.; Cosenza, M.; Fragai, M.; Lee, Y.-M.; Luchinat, C.; Mangani, S.; Terni, B.; Turano, P. *Proc. Natl. Acad. Sci. USA.* **2005**, *102* (15), 5334.

(72) Galili, U.; Shohet, S. B.; Kobrin, E.; Stults, C. L.; Macher, B. A. *J. Biol. Chem.* **1988**, *263* (33), 17755.

(73) Galili, U.; Rachmilewitz, E. A.; Peleg, A.; Flechner, I. *J. Exp. Med.* **1984**, *160* (5), 1519.

(74) Joziasse, D. H.; Oriol, R. *BBA-Mol Basis Dis.* **1999**, 1455 (2-3), 403.

(75) Galili, U. *Immunol. Cell Biol.* **2005**, *83* (6), 674.

(76) Galili, U.; Minanov, O. P.; MlchSer, R. E.; Stone, K. R. *Xenotransplantation* **1997**, *4* (3), 127.

(77) Chen, G.; Qian, H.; Starzl, T.; Sun, H.; Garcia, B.; Wang, X.; Wise, Y.; Liu, Y.; Xiang, Y.; Copeman, L.; Liu, W.; Jevnikar, A.; Wall, W.; Cooper, D. K. C.; Murase, N.; Dai, Y.; Wang, W.; Xiong, Y.; White, D. J.; Zhong, R. *Nat. Med.* **2005**, *11* (12), 1295.

(78) Yang, Y.-G.; Sykes, M. *Nat. Rev. Immunol.* **2007**, *7* (7), 519.

(79) Schuurman, H.-J.; Cheng, J.; Lam, T. *Xenotransplantation* **2003**, *10* (4), 293.

(80) Cozzi, E.; Bosio, E.; Seveso, M.; Vadori, M.; Ancona, E. *Br. Med. Bull.* **2005**, *75–76*, 99.

(81) Welsh, R. M.; O'Donnell, C. L.; Reed, D. J.; Rother, R. P. *J. Virol.* **1998**, *72* (6), 4650.

(82) Huai, G.; Qi, P.; Yang, H.; Wang, Y. *Int. J. Mol. Med*. **2016**, 37 (1), 11.

(83) Kitajima, K.; Varki, N.; Sato, C. In *Sialoglyco chemistry and biology ii: tools and techniques to identify and capture sialoglycans*; GerardySchahn, R and Delannoy, P and VonItzstein, M., Ed.; Topics in Current Chemistry; Springer Int Publishing AG, **2015**, 367, 75.

(84) Gao, H.; Li, S.; Sun, W. Q.; Yun, Z.; Zhang, X.; Song, J.-W.; Zhang, S.-K.; Leng, L.; Ji, S.-P.; Tan, Y.-X.; Gong, F. *Tissue Eng. Part C. Methods* **2015**, *21* (11), 1197.

(85) Jia, N.; Barclay, W. S.; Roberts, K.; Yen, H. L.; Chan, R. W. Y.; Lam, A. K. Y.; Air, G.; Peiris, J. S. M.; Dell, A.; Nicholls, J. M.; Haslam, S. M. *J. Biol. Chem.* **2014**, *289* (41), 28489.

(86) Hakomori, S. *Pure Appl. Chem.* **1991**, *63* (4), 473.

(87) Saeland, E.; van Vliet, S. J.; Bäckström, M.; van den Berg, V. C. M.; Geijtenbeek, T. B. H.;

Meijer, G. A.; van Kooyk, Y. *Cancer Immunol. Immun.* **2007**, *56* (8), 1225.
(88) Angata, T.; Varki, A. *Chem. Rev.* **2002**, *102* (2), 439.
(89) Varki, A. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **1997**, *11* (4), 248.
(90) Varki, A.; Schauer, R. *Essentials Glycobiol. - Second Ed.* **2014**, *9*, 1.
(91) Varki, A.; Hooshmand, F.; Diaz, S.; Varki, N. M.; Hedrick, S. M. *Cell* **1991**, *65* (1), 65.
(92) Mandal, C.; Schwartz-Albiez, R.; Vlasak, R. Gerardy-Schahn, R., Delannoy, P., von Itzstein, M., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, **2015**; 1.
(93) Gagneux, P.; Cheriyan, M.; Hurtado-Ziola, N.; van der Linden, E. C.; Anderson, D.; McClure, H.; Varki, A.; Varki, N. M. *J Biol Chem* **2003**, *278*.
(94) Nicholls, J. M.; Bourne, A. J.; Chen, H.; Guan, Y.; Peiris, J. S. M. *Respir. Res.* **2007**, *8* (1), 73.
(95) Davies, L. R. L.; Varki, A. Gerardy-Schahn, R., Delannoy, P., von Itzstein, M., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, **2015**; pp 31–54.
(96) Taylor, R. E.; Gregg, C. J.; Padler-Karavani, V.; Ghaderi, D.; Yu, H.; Huang, S.; Sorensen, R. U.; Chen, X.; Inostroza, J.; Nizet, V.; Varki, A. *J. Exp. Med.* **2010**, *207* (8), 1637 LP-1646.
(97) van Riel, D.; Munster, V. J.; de Wit, E.; Rimmelzwaan, G. F.; Fouchier, R. A.; Osterhaus, A. D.; Kuiken, T. *Am. J. Pathol.* **2007**, *171*, 1215.
(98) Shinya, K.; Ebina, M.; Yamada, S.; Ono, M.; Kasai, N.; Kawaoka, Y. *Nature* **2006**, *440* (7083), 435.
(99) Baigent, S. J.; McCauley, J. W. *BioEssays* **2003**, *25* (7), 657.
(100) Sakai, T.; Nishimura, S. I.; Naito, T.; Saito, M.; Melancon, S. B. *Sci. Rep.* **2017**, *7* (March), 45043.
(101) Fouchier, R. A. M.; Munster, V.; Wallensten, A.; Bestebroer, T. M.; Herfst, S.; Smith, D.; Rimmelzwaan, G. F.; Olsen, B.; Osterhaus, A. D. M. E. *J. Virol.* **2005**, *79* (5), 2814.
(102) Webster, R. G.; Bean, W. J.; Gorman, O. T.; Chambers, T. M.; Kawaoka, Y. *Microbiol. Rev.* **1992**, *56* (1), 152.
(103) Salomon, R.; Webster, R. G. *Cell* **2009**, *136* (3), 402.
(104) Peiris, J. S. M.; de Jong, M. D.; Guan, Y. *Clin. Microbiol. Rev.* **2007**, *20* (2), 243.
(105) Paules, C.; Subbarao, K. *Lancet* **2017**, *390* (10095), 697.
(106) Ibricevic, A.; Pekosz, A.; Walter, M. J.; Newby, C.; Battaile, J. T.; Brown, E. G.; Holtzman, M. J.; Brody, S. L. *J. Virol.* **2006**, *80*, 7469.
(107) Russell, R. J.; Stevens, D. J.; Haire, L. F.; Gamblin, S. J.; Skehel, J. J. *Glycoconj. J.* **2006**, *23*, 85.
(108) Shinya, K.; Ebina, M.; Yamada, S.; Ono, M.; Kasai, N.; Kawaoka, Y. *Nature* **2006**, *440*, 435.
(109) Skehel, J. J.; Wiley, D. C. *Annu. Rev. Biochem.* **2000**, *69*.
(110) Connor, R. J.; Kawaoka, Y.; Webster, R. G.; Paulson, J. C. *Virology* **1994**, *205* (1), 17.
(111) Glaser, L.; Zamarin, D.; Acland, H. M.; Spackman, E.; Palese, P.; García-Sastre, A.; Tewari, D. *Glycoconj. J.* **2006**, *23* (1), 93.
(112) Kumari, K.; Gulati, S.; Smith, D. F.; Gulati, U.; Cummings, R. D.; Air, G. M. *Virol. J.* **2007**, *4*, 42.
(113) Matrosovich, M.; Matrosovich, T.; Uhlendorff, J.; Garten, W.; Klenk, H.-D. *Virology* **2007**, *361* (2), 384.
(114) Matrosovich, M. N.; Matrosovich, T. Y.; Gray, T.; Roberts, N. A.; Klenk, H. D. *Proc. Natl. Acad. Sci. USA.* **2004**, *101*, 4620.

(115) Tumpey, T. M.; Maines, T. R.; Van Hoeven, N.; Glaser, L.; Solórzano, A.; Pappas, C.; Cox, N. J.; Swayne, D. E.; Palese, P.; Katz, J. M.; García-Sastre, A. *Science (80-.).* **2007**, *315* (5812), 655 LP-659.

(116) Nycholat, C. M.; McBride, R.; Ekiert, D. C.; Xu, R.; Rangarajan, J.; Peng, W.; Razi, N.; Gilbert, M.; Wakarchuk, W.; Wilson, I. A.; Paulson, J. C. *Angew. Chem.- Int. Ed.* **2012**, *51* (20), 4860.

(117) Macchi, E.; Rudd, T. R.; Raman, R.; Sasisekharan, R.; Yates, E. A.; Naggi, A.; Guerrini, M.; Elli, S. *Biochemistry* **2016**, *55* (48), 6605.

(118) Zhang, W.; Shi, Y.; Qi, J.; Gao, F.; Li, Q.; Fan, Z.; Yan, J.; Gao, G. F. *J. Virol.* **2013**, *87* (10), 5949.

(119) Woods Group. (2005-2017) GLYCAM Web. Complex Carbohydrate Research Center, University of Georgia, Athens, GA. (http://glycam.org).

(120) Shriver, Z.; Raman, R.; Viswanathan, K.; Sasisekharan, R. *Chem. Biol.* **2009**, *16* (8), 803.

(121) Chandrasekaran, A.; Srinivasan, A.; Raman, R.; Viswanathan, K.; Raguram, S.; Tumpey, T. M.; Sasisekharan, V.; Sasisekharan, R. *Nat Biotechnol* **2008**, *26* (1), 107.

(122) Sassaki, G. L.; Elli, S.; Rudd, T. R.; Macchi, E.; Yates, E. A.; Naggi, A.; Shriver, Z.; Raman, R.; Sasisekharan, R.; Torri, G.; Guerrini, M. *Biochemistry* **2013**, *52* (41), 7217.

(123) Blixt, O.; Brown, J.; Schur, M. J.; Wakarchuk, W.; Paulson, J. C. *J. Org. Chem.* **2001**, *66* (7), 2442.

(124) Logan, S. M.; Altman, E.; Mykytczuk, O.; Brisson, J.-R.; Chandan, V.; Michael, F. St.; Masson, A.; Leclerc, S.; Hiratsuka, K.; Smirnova, N.; Li, J.; Wu, Y.; Wakarchuk, W. W. *Glycobiology* **2005**, *15* (7), 721.

(125) Unverzagt, C.; Kelm, S.; Paulson, J. C. *Carbohydr. Res.* **1994**, *251* (C), 285–301.

(126) Yu, H.; Chokhawala, H.; Karpel, R.; Yu, H.; Wu, B.; Zhang, J.; Zhang, Y.; Jia, Q.; Chen, X. *J. Am. Chem. Soc.* **2005**, *127* (50), 17618.

(127) Weinstein, J.; de Souza-e-silva, U.; Paulson, J. C. *J. Biol. Chem.* **1982**, *257* (22), 13845.

(128) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622.

Chapter 4

NMR applications to dynamic and other systems

4.1. Introduction

Proteins are inherently dynamic and sample a vast ensemble of conformations. Thus, protein conformations, different than the native one, may play an important role in molecular recognition.[1] In this context, NMR has permitted the study of the dynamics of globular and structural well defined proteins as well as the structural characterization of disordered ones, thus overcoming the limitations of X-ray crystallography.

Not only protein dynamics play a key role in the molecular recognition event. The conformation and dynamics of the ligands are also rather important to explain the binding events. Therefore, the conformational analysis of the bound ligand is also essential to understand the binding process.

In summary, molecular recognition processes involve dynamic players that adapt their 3D arrangement to favor the binding. Only under certain conformational conditions, the interaction between partners occurs. In this context, NMR experimental data are essential to get insights into the conformational behavior of ligands and receptors in their free and bound states.

NMR can also be employed for the real-time monitoring of the evolution of chemical reactions. The gained information can be employed to mechanistically understand the transformation of reactants into final products. Moreover, NMR has also provided key insights in the deconvolution of the interaction of dynamic combinatorial libraries with receptors, a process where multiple reversible reactions takes place.

Since the conceptual NMR-based approaches to study sugar interactions are general within the molecular recognition field, I have also employed analogous methodologies to study other ligand-receptor interaction processes related to interesting biological and biomedical events, in collaboration with other research groups. In particular, in the following sections of this chapter, I will describe different examples of the application of NMR methods to the study of the structural, conformational and dynamics features of different systems with diverse chemical nature: from the analysis of the degree of ordering of flexible protein domains, to the characterization of the interaction of epothilone derivatives with tubulin, including the evaluation of the recognition of a family of molecules by the DNA minor groove. Finally, the possible applications of NMR in dynamic combinatorial chemistry have also been scrutinized.

4.2. NMR structural analysis of the N-terminal region of the C/EBP homologous protein (N-CHOP)

It is highly recognized that protein secondary and tertiary structures are essential for protein functions. However, it is now clear that disordered and unfolded protein domains can also be biologically active following an external stimulus.[2,3] The concept "intrinsically disordered protein" (IDP) refers to proteins without a defined tertiary structure at physiological conditions. The genomic analysis of IDPs indicates that their proportion increase with the organism complexity.[4–7] However, IDPs often display evidences of local order, which may be crucial for their interactions with other partners.[8–12] In fact, the recognition of an external stimulus by the IDP generally triggers the adoption of a well-defined tertiary structure. The response to a specific stimulus is usually related with the cellular location in which the interaction takes place. Thus, IDPs act as occasional workers at punctual locations in the cell.

Right now, most of the characterized IDPs are associated to cell cycle control and DNA regulation[6,11,13–16] However, IDPs can also be found as independent proteins or as domains of partially structured proteins.[2,5]

Even though X-ray crystallography methods cannot be employed to characterize the structure of IDPs due to its high conformational heterogeneity, fortunately a variety of NMR methods are suitable for the structural and dynamic elucidation of IDPs. In this manner, the folding propensities of different IDPs have been unraveled by analyzing their heteronuclear ^{1}H-^{15}N NMR spectra, those that permit measuring the longitudinal (R_1) and transversal (R_2) relaxation rates as well as the steady-state heteronuclear ^{1}H-^{15}N NOE (hetNOE). These parameters display high sensitivity to the adoption of ordered regions on IDPs within the ps-ns time scale.[17,18]

Protein dynamics may be defined by measuring the motional freedom of the NH vector of each amino acid residue and defining the associated rotational motion correlation time (τ_c). The amino acids belonging to a structured protein domain have a restricted motion due to the adoption of a stable 3D structure. In this case, they display a large τ_c. In contrast, amino acids that lack a stable secondary and/or tertiary structure display short τ_c. The hetNOE may also be employed to discriminate well-structured from flexible regions, due to its change in sign (figure 1).

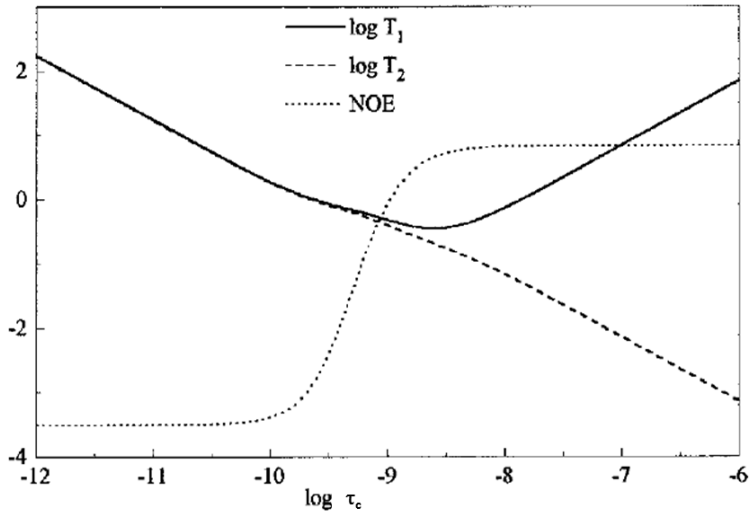


Figure 1. Representation of the T_1, T_2 and hetNOE against the total rotational motion correlation time (τ_c).

Herein we discuss the conformational features as IDP of the N-terminal domain of C/EBP homologous protein (CHOP) protein, a partially unstructured protein of the family of CCAAT/enhancer-binding protein (C/EBP). The C/EBP family protein consists on a group of proteins present in the cytosol of the cells that regulates DNA transcription[19] and has shown an apoptotic response to DNA damage or stress of the endoplasmic reticulum.[20] The C/EBP family has a highly conserved bzip ("basic leucine zipper") domain that binds to a consensus sequence of nucleotides while the N-terminus is diverse among the family and regulates the transcription activity of CHOP.[19] The bzip leucine zipper is expected to adopt a α-helix structure when bound to DNA, whereas the N-terminus domain behaves as an IDP (figure 2).[21] Because the structural studies of the CHOP are scarce and CHOP is of interest due to its implications in diabetes or neurodegenerative pathologies,[22,23] the elucidation of the dynamic features of CHOP may shed light on the regulation of the activity of CHOP, a potential therapeutic target.

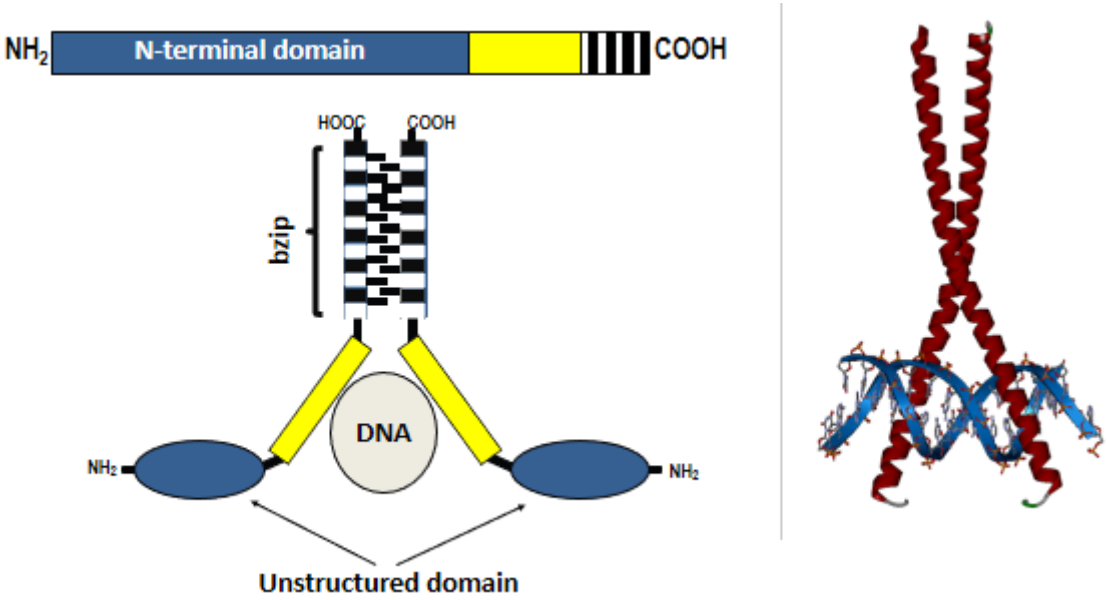


Figure 2. On the left: scheme of CHOP protein. On the right: Model of interaction of CHOP with DNA.

4.2.1. Results and discussion

The backbone dynamics of CHOP were analyzed by NMR measuring its relaxation parameters on a ^{15}N labeled domain provided by Dr. Carlos Fernández Tornero at CIB-CSIC.

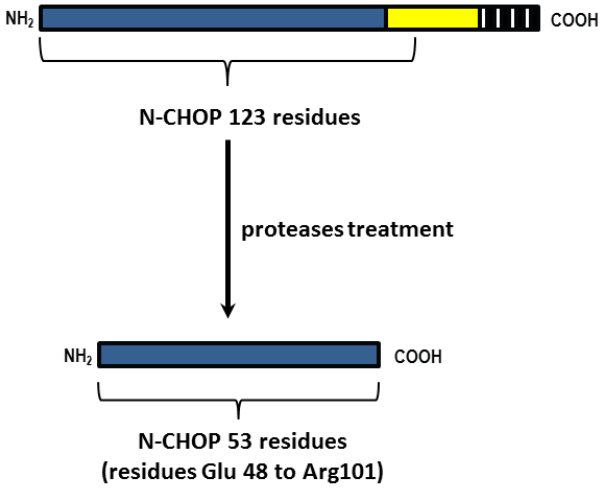


Figure 3. N-terminal domain fragment studied by NMR.

Prior to this analysis, CHOP N-terminus domain fragment (N-CHOP, 53 amino acids from Glu 48 to Arg 93) (figure 3) assignment was achieved employing direct ^{13}C-NMR detection experiments (CBCANCO, CON) (figure 4). This approach was employed since IDPs display a lack of ^{1}H signal dispersion (figure 5 and 6) and fast HN-water exchange precludes the assignment by traditional

^{15}N-based methods. A previous chemical-shift index analysis[24] of the chemical shifts of Cα and Cβ had predicted the existence of a α-helix structure propensity comprising amino acids 60-68. To confirm the structural results, the dynamics of N-CHOP, containing this region, was evaluated by measuring hetNOEs, R_1 and R_2 rates.

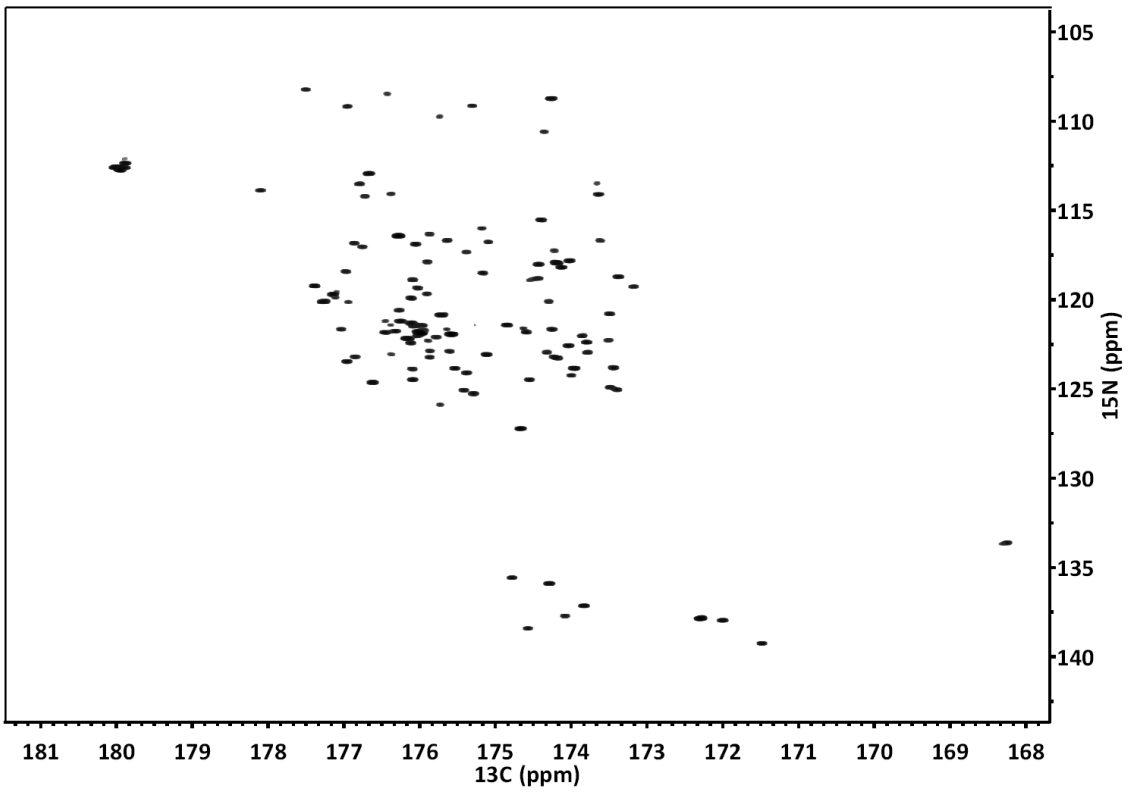


Figure 4. The CON NMR spectrum of full N-terminal domain of CHOP (700 MHz, 298 K).

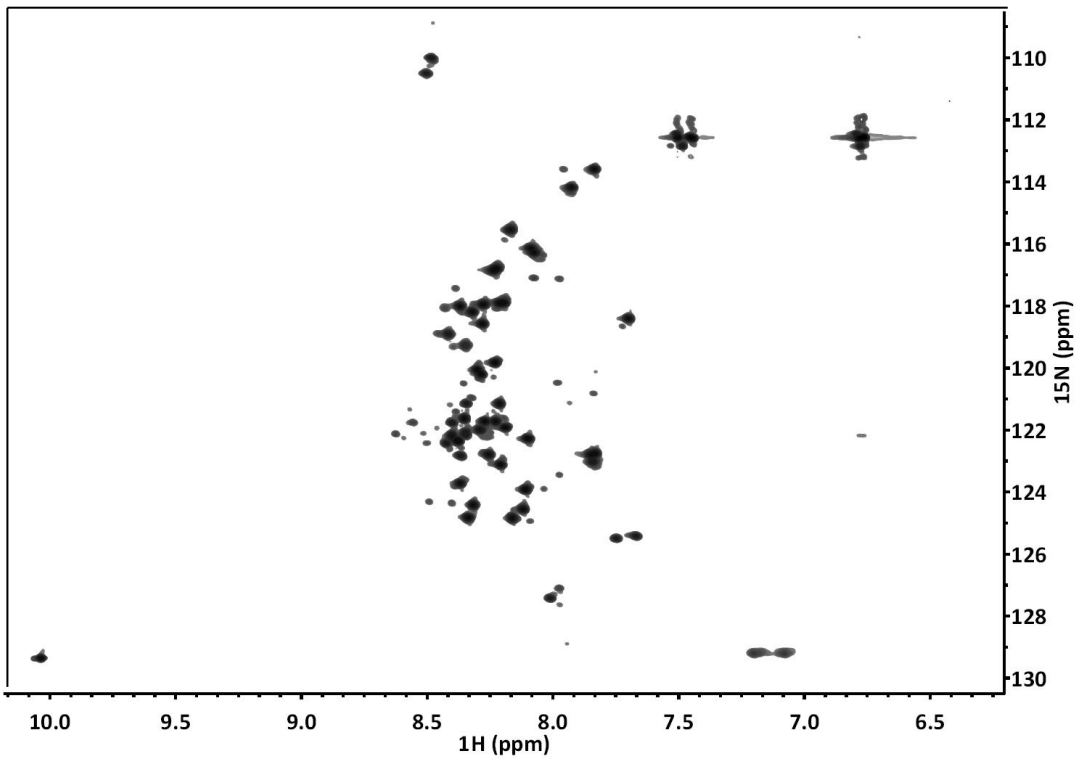


Figure 5. The ^{1}H-^{15}N HSQC spectrum ^{15}N labelled full N-terminal domain of CHOP (1 mM, 700 MHz, 298 K). The signals of the NH of the backbone appear in a very narrow region (δ 7.5-8.5 ppm) typically observed for unstructured proteins.

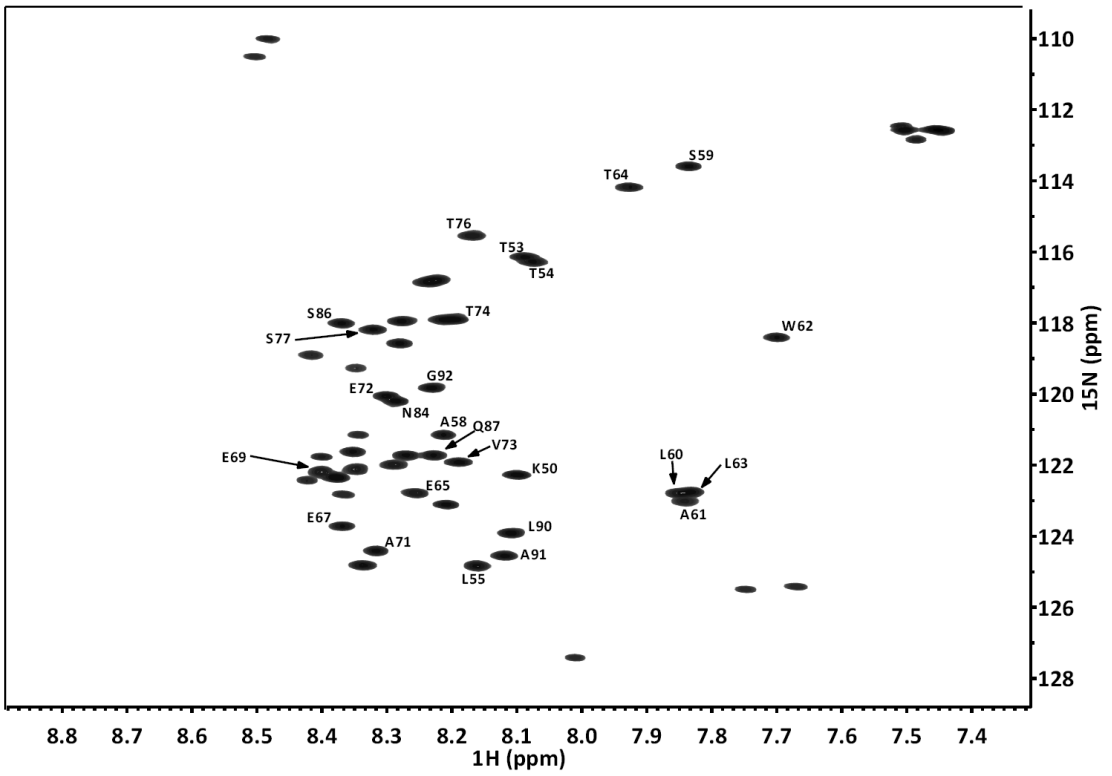


Figure 6. The ^{1}H-^{15}N HSQC of N-terminal domain fragment of CHOP (N-CHOP, 53 amino acids: Glu48 to Asp101) (700 MHz, 298 K).

For anisotropic systems as IDPs, R_1 ($1/T_1$) and R_2 ($1/T_2$) analysis is restricted by the model-free approach[25] analysis, whilst hetNOEs (figure 7) are sensitive reporters of the presence of fast motions and provides information on the regions with different mobility.[17]

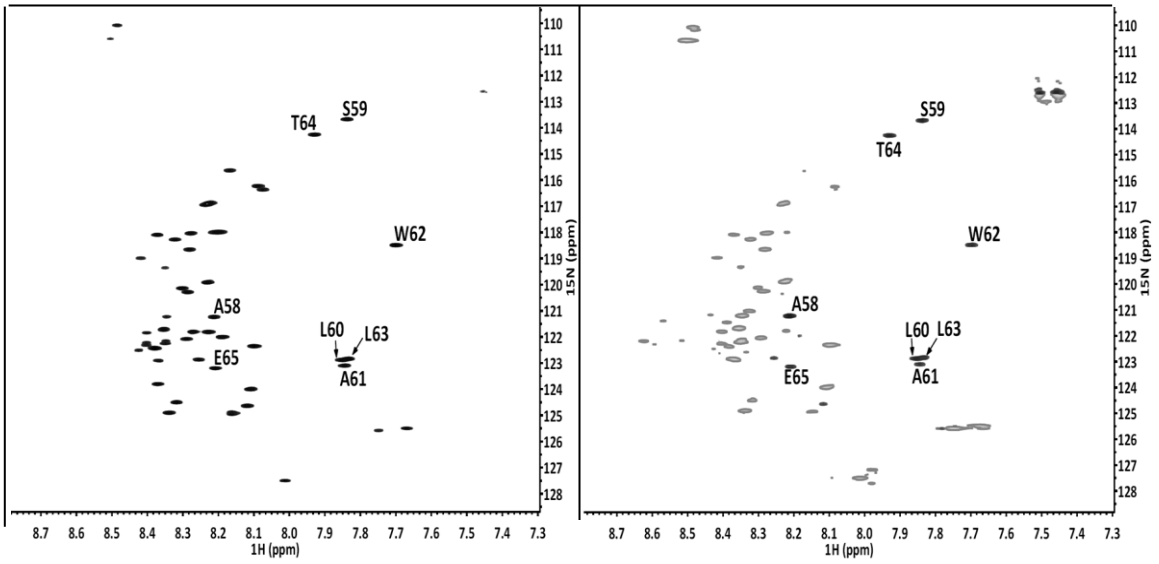


Figure 7. The ^{1}H-^{15}N hetNOE spectra of N-CHOP (700 MHz, 298 K). Left panel: ^{1}H-^{15}N HSQC of N-CHOP at equilibrium, without proton saturation. Right panel: ^{1}H-^{15}N HSQC of N-CHOP with proton saturation. Peaks corresponding to amino acids with positive hetNOE are annotated.

The analysis of the hetNOE NMR spectra (figures 7 and 8) revealed the existence of a region with positive hetNOE values that comprises the amino acids from serine 59 to glutamine 65. They display a large τ_c, characteristic of structured amino acids, strongly suggesting that this region (Ser59 to Glu65) has a propensity to adopt secondary structure. In contrast, the rest of the backbone NHs displayed negative hetNOE value, associated with short τ_c and thus, to the presence of fast and mobile amino acids, as frequently occurs at the N-terminal and C-terminal ends, in many occasions behaving as disordered regions.

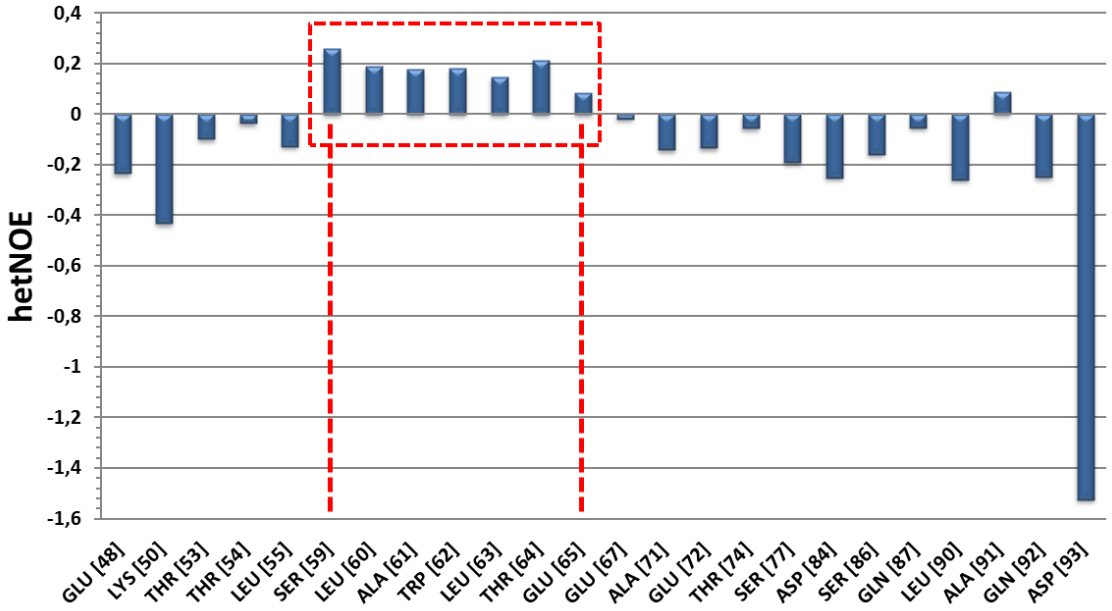


Figure 8. Histogram of the ^{1}H-^{15}N hetNOE analysis of N-CHOP. The change in sign in the Ser59 to Glu65 region is evident.

As mentioned above (figure 1), the evaluation of the relaxation rates R_1 and R_2 for anisotropic models (i.e. IDPs) helps to detect structured regions. Typically, structured macromolecules display a large τ_c with short R_1 (large T_1) and large R_2 (short T_2). On the other hand, small molecules and unstructured domains display short τ_c with short R_1 and R_2.

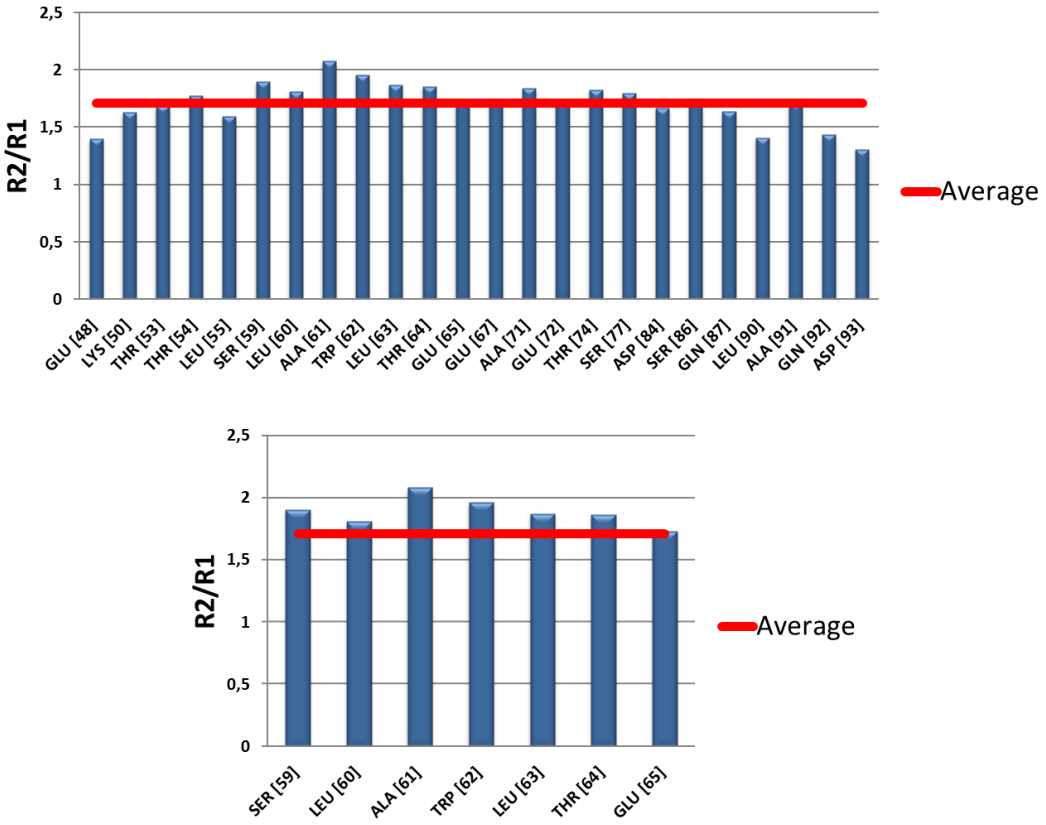


Figure 9. Histogram of the ratio R_2/R_1 of N-CHOP.

The R_2/R_1 ratio for N-CHOP (figure 9) displays larger values for Ser59 to Glu65 region, while the rest of the amino acids, in particular those close to the N- and C-terminus, are more mobile.

NMR methods also provide information on the chemical exchange process between the water from the solvent and the exchangeable backbone NHs whose exchange rate will depend of their solvent accessibility and hence of the structure. The ^{1}H chemical exchange with the solvent is usually significantly restricted in the case of the non-accessible NHs of the backbone in structured domains. In contrast, flexible and unstructured domains are exposed to the solvent, thus facilitating the chemical exchange. Thus, chemical exchange-phase modulated experiments (CLEANEX-PM) were acquired to observe the chemical exchange of the NHs with water from the solvent based in the magnetization transfer from the selectively excited solvent protons to the exchangeable protons of the protein backbone. The advantage of CLEANEX-PM experiments resides in their ability to suppress the NOE contributions from Cα Hs but also the exchange-relayed NOE peaks, only displaying those peaks affected by the chemical exchange process.[26,27]

CLEANEX-PM filtered experiments were recorded at different mixing times (figure 10). At increasing mixing times, peak intensities gradually increase as function of the proton exchange rate with solvent. The analysis of these experiments revealed that the amino acids of the key region (Ser59 to Glu68) have a reduced exchange rate or even do not appear in the spectrum, thus confirming that they are less exposed to the solvent.

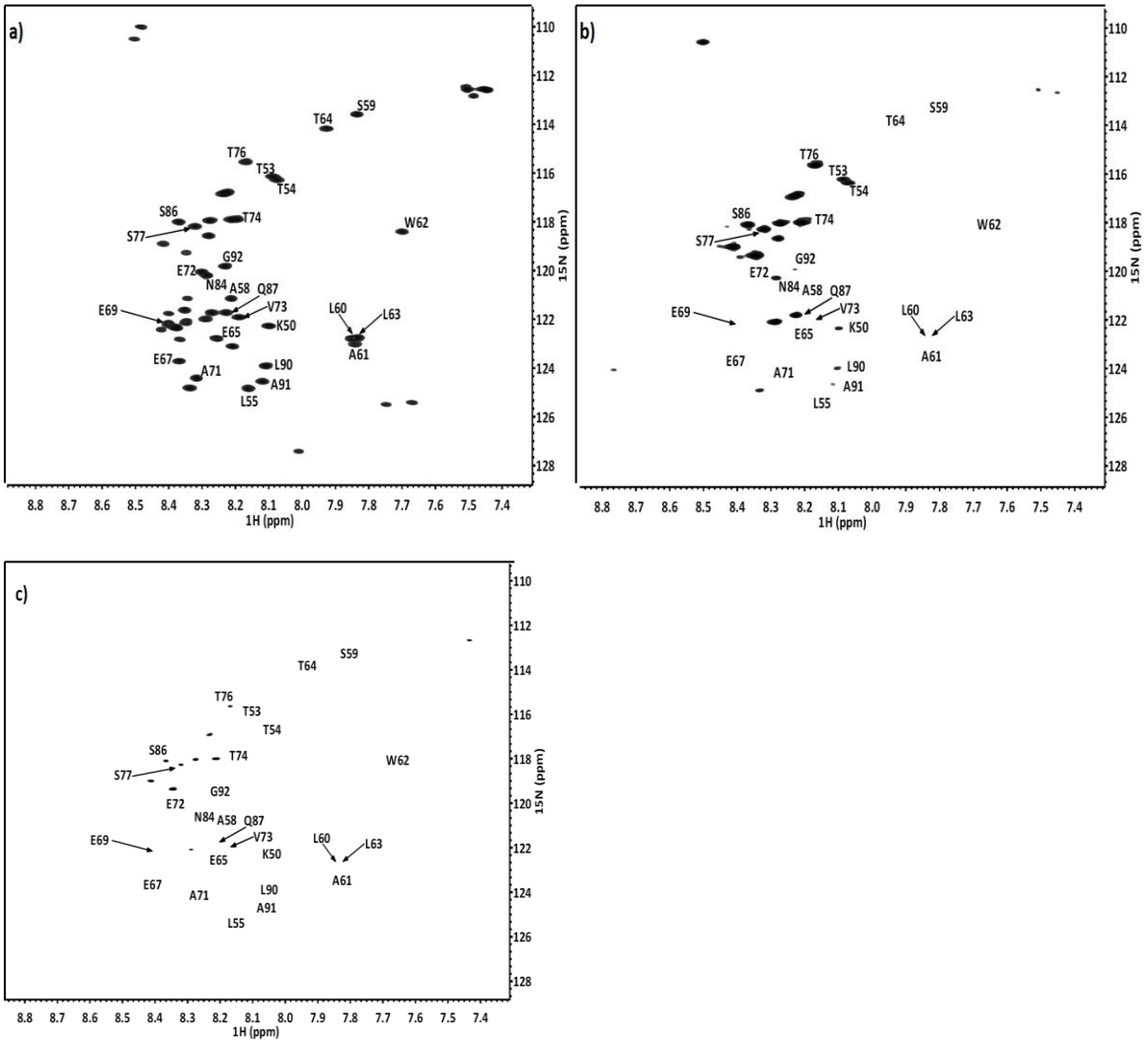


Figure 10. CLEANEX-PM spectra of N-CHOP (700 MHz, 298 K). Panel A: ^{1}H-^{15}N HSQC spectrum. Panel B: CLEANEX-PM spectrum with 24 ms of mixing time. Panel C: CLEANEX-PM spectrum with 6 ms of mixing time.

4.2.2. Conclusions

The NMR-based dynamic analysis of the N-terminus domain of CHOP has revealed evidences of a propensity to adopt a defined secondary structure in the Ser59 to Glu65 region. The measured dynamic parameters, positive hetNOE and large R_2/R_1 ratios, are associated to large τ_c values characteristic of structured proteins. Furthermore, ^{1}H chemical exchange experiments have also shown that the amino acids of this region have a significantly reduced chemical exchange with water solvent. This fact suggests that they are protected from the solvent. Based on a previous chemical shift index analysis, this region may adopt a α-helix secondary structure.

4.2.3. Experimental

R_1, R_2, and ^1H-^{15}N hetNOE experiments were acquired on a 1 mM sample of ^{15}N labelled of N-CHOP derived from proteolysis of the full protein. The fragment of 53 amino acids was provided by Dr. Carlos Fernández Tornero (CIB-CSIC). A SE buffer containing 90 % H_2O and 10 % D_2O was employed.

R_1, R_2, and ^1H-^{15}N hetNOE experiments targeting the ^{15}N spins of the backbone and tryptophan side chain were performed at 700 MHz Bruker Avance spectrometer at CERM (Florence) by Dr. Isabella Felli. Experimentally, for relaxation measurements, 10 data points were acquired interleaved with relaxation delays in the range 20 ms - 1.5 s (20, 100, 200, 350, 500, 600, 800, 1000, 1200 and 1500 ms) for R_1 and 31.36 ms - 0.47 s (31.36, 62.72, 94.08, 125.44, 156.80, 219.52, 283, 344.96, 407.68 and 470.40 ms) for R_2, using a delay between experiments of 3 s. R_2 was measured using a 0.9 ms delay between refocusing pulses in the CPMG train. Heteronuclear ^1H-^{15}N NOE was measured in an interleaved manner using a ^1H saturation time of 3 s and a relaxation delay of 3 s. For the saturation, 120° pulses were used. CLEANEX-PM filtered experiments were acquired at different mixing times for CLEANEX spin lock module[26] (6, 12, 18 and 24 ms). All experiments were acquired at 298 K.

- Data analysis

The hetNOE, R_1 and R_2 data were deduced from the NMR spectra with Bruker software using the peak intensity values for the calculations. R_1 and R_2 values were obtained by fitting the decay of each peak intensity to an exponential function as implemented in the software Dynamic Center.

The ^1H-^{15}N assignment was translated from the previous assignment provided by Dr. Ángeles Canales (Complutense University of Madrid).

4.3. Microtubule stabilizing agents: The binding of Epothilone analogues to tubulin

Epothilones (figure 11) are a group of molecules with a common 16-membered ring that are potent molecules to stabilize microtubules (microtubule-stabilizing agents, MSA) by interacting with tubulin. In this manner, they arrest the cell cycle resulting in apoptosis.[28–30] Indeed, epothilones have been used as potent anti-tumor drugs in chemotherapy.[31]

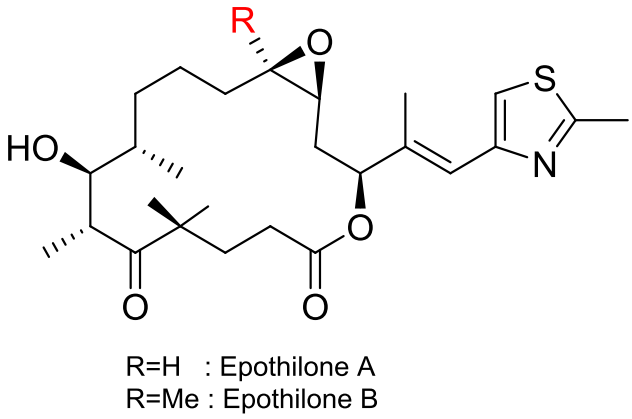


Figure 11. Scheme of the structure of epothilone A and B.

Two binding sites coexist in tubulin that may be involved in epothilone binding, dubbed the luminal site[32,33] and pore site.[34,35] The luminal site is an internal site of the β-subunit, while the pore site is external, only present in microtubules, and involves the α and β subunits of different heterodimers.

Given the favorable biopharmaceutical properties of epothilones, pharmacodynamics and pharmacokinetics, different derivatives have been synthesized to further improve their (already good) properties. In fact, several epothilone derivatives have gone into advance stages of clinical trials[36–39] and some of them have already been approved by the FDA.[40] Some of the synthetic derivatives have replaced the epoxide moiety by a cyclopropane.[41–43] In fact, the epothilone derivative studied in this section (figure 12) consists in a trans-configured cyclopropane derivative of the epothilone A (EPO-tc) that has been shown to display a strong activity in in-vitro tests using a wide sample of cell lines.

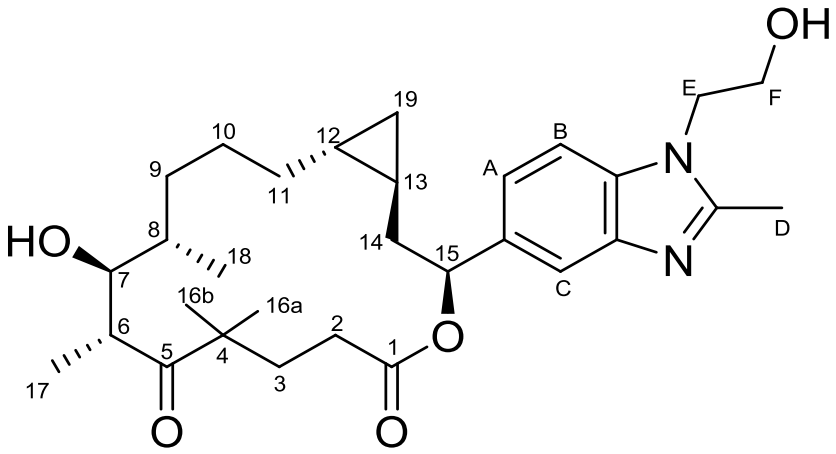


Figure 12. Scheme of the trans-cyclopropane epothilone (EPO-tc) derivative studied in this section.

4.3.1. Results and discussion

After assignment of the ^{1}H NMR resonances of EPO-tc using standard NMR methods, its interaction with tubulin was addressed from the ligand point of view. Thus, STD-NMR experiments were performed for EPO-tc in the presence of dimeric tubulin and microtubules. Dimeric tubulin only displays the luminal binding site while the pore site only appears in microtubules.

The STD spectrum of EPO-tc in presence of microtubules (figure 13) showed that every EPO-tc proton receives saturation from the protein in a high extent (table 1). Unfortunately, the presence of the GMPCPP6 nucleotide analogue to stabilize the sample precludes the accurate quantification of several EPO-tc signals.

When the STD-NMR experiment was performed in the presence of dimeric tubulin (figure 14), information on the binding at the luminal site of the β-subunit was extracted.[44,45] In both cases, the derived epitope of EPO-tc (figure 15) indicates that the epothilone ring as well as the aromatic region participates in the binding event. The derived epitopes in both cases (figure 16) followed a similar pattern supporting the binding towards the luminal site. However, since in the presence of dimeric tubulin fewer NMR signals could be identified in a non-ambiguous manner, this conclusion should be considered with caution.

A conformational analysis protocol was then performed to interpret the STD data and to fully describe the binding event. Thus, tr-NOESY experiments were performed to get information on the interproton distances in the bound state and thus, on the 3D arrangement of the ligand.

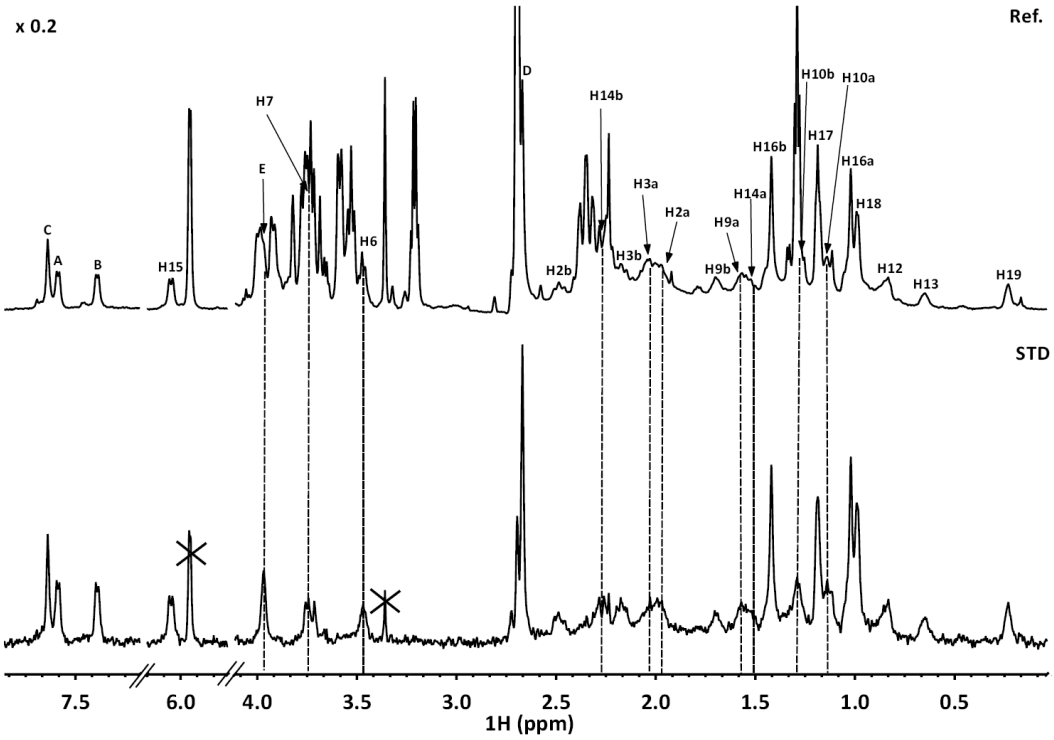


Figure 13. The STD (below) and off-resonance (on top) NMR spectra (600 MHz, cryoprobe) of a microtubule:EPO-tc mixture (1:45 molar ratio, 10 μM microtubule, 0.1 mM GMPCPP6, 450 μM EPO-tc). The temperature was 310 K and a 2 seconds saturation at δ -1.5 ppm was employed. A spin lock filter was used to minimize the background of the protein. Signals with X correspond to stabilizer nucleotide and buffer impurities.

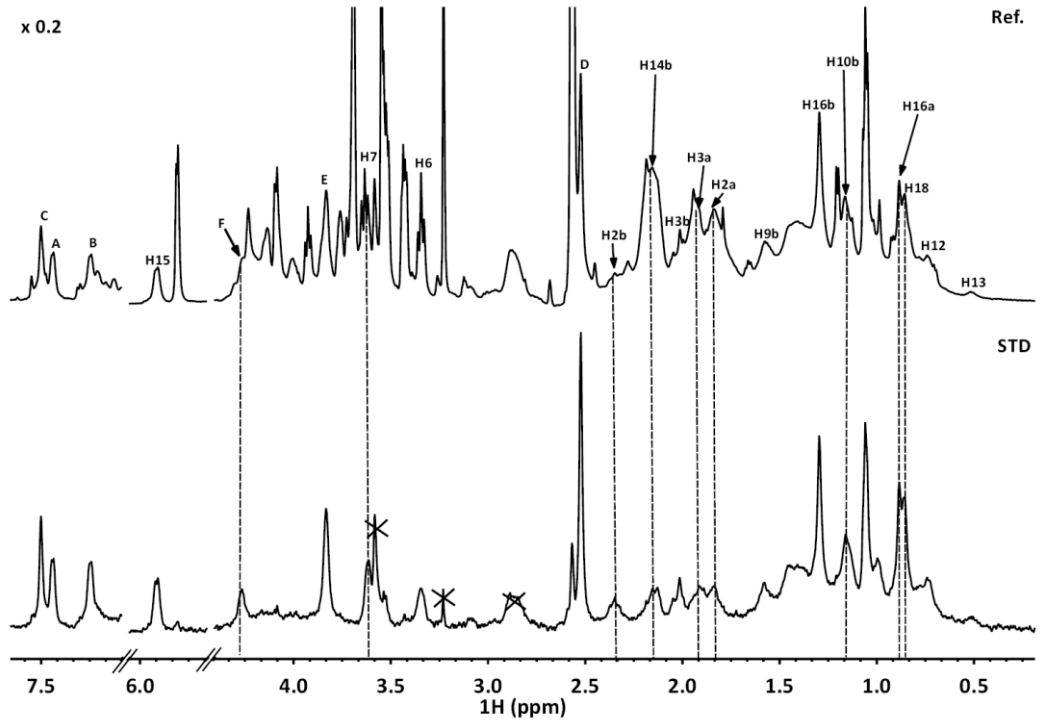


Figure 14. STD (below) and off-resonance (on top) NMR spectra (600 MHz, cryoprobe) of a dimeric tubulin:EPO-tc mixture (1:45 molar ratio, 10 μM dimeric tubulin). The temperature was 298 K and a 2 seconds saturation at δ -1.5 ppm was employed. A spin lock filter was used to minimize the background of the protein protons. Signals with X correspond to stabilizer nucleotide and buffer impurities.

	Microtubules		Dimeric tubulin	
	Absolute STD (%)	Relative STD (%)	Absolute STD (%)	Relative STD (%)
H2a	22.3	70	-	-
H2b	22.3	70	13.8	46
H3a	22.3	70	-	-
H3b	30.4	95	26.6	89
H6	13.0	41	n.d.	n.d.
H7	n.d.	n.d.	n.d.	n.d.
H8	n.d.	n.d.	-	-
H9a	28.7	90	-	-
H9b	23.8	75	15.0	50
H10a	n.d.	n.d.	-	-
H10b	n.d.	n.d.	17.8	59
H11a	n.d.	n.d.	-	-
H11b	n.d.	n.d.	-	-
H12	27.1	85	23.0	76
H13	31.9	100	23.0	76
H14a	28.7	90	-	-
H14b	n.d.	n.d.	n.d.	n.d.
H15	30.7	96	29.1	97
H16a	27.6	86	-	-
H16b	24.3	76	20.4	68
H17	25.8	81	-	-
H18	31.3	98	-	-
H19	31.9	100	-	-
A	31.9	100	29.1	97
B	31.9	100	29.1	97
C	31.0	97	30.0	100
D	25.6	80	26.6	89
E	21.6	68	21.9	73
F	-	-	18.4	61

Table 1. STD intensities of the proton signals of EPO-tc. n.d.: not determined (signals were observed but quantification is not accurate due to overlapping)

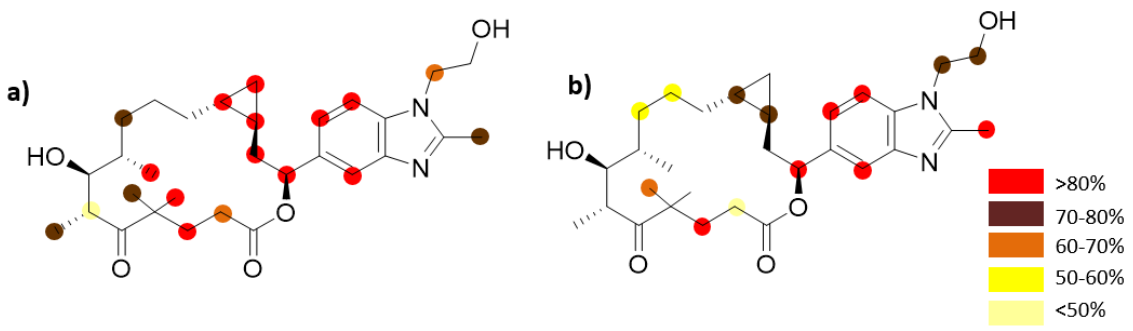


Figure 15. Epitope mapping as deduced by STD-NMR for EPO-tc. Panel a: the binding epitope to microtubules. Panel B: the binding epitope to dimeric tubulin.

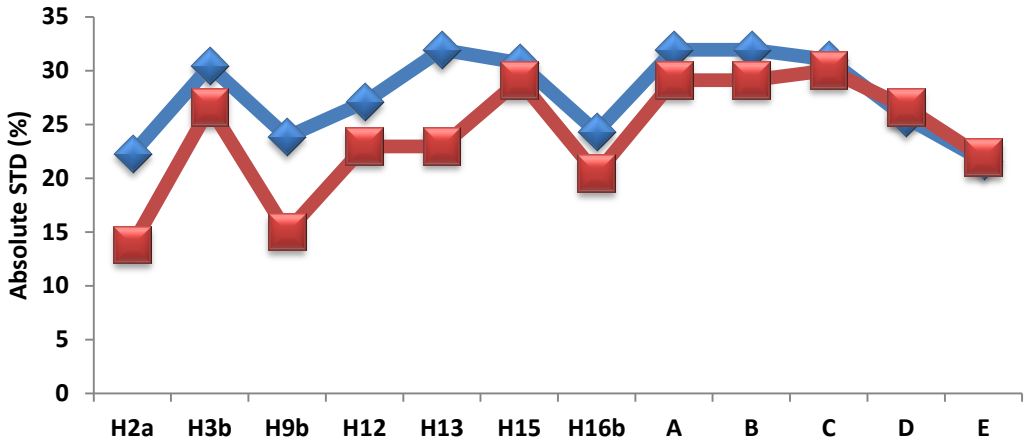


Figure 16. Observed STD intensities for the different protons of EPO-tc with the different tubulins. The data are given for microtubules (in blue) and for dimeric tubulin (in red).

Interproton NOE	Intensity
H2-CH_3 18	Weak
H2-H12	Medium
H3-H12	Strong
H6- H12	Weak
H6- CH_3 17	Strong
H6- CH_3 16	Strong
H7-CH_3 18	Strong
H7-CH_3 17	Strong
H14-H12	Strong
C-H15	Medium
C-H3	Weak
C-H12	Weak
E-F	Strong
E-D	Strong
CH_3 17- CH_3 18	Weak

Table 2. The observed NOEs in the tr-NOESY experiment perfomed for EPO-tc in the presence of microtubules.

The combined analysis of NOE data (table 2) together with a conformational search protocol allowed to deduce the bound conformation. The torsion angles that defined the EPO-tc ring are shown in table 3 in comparison with the reference structure of EPO-A reported by Canales *et al*.[46]

Interestingly, the only significant differences are located in the bonds close to the trans-cyclopropane ring of EPO-tc in accordance with a necessary adaptation to the binding site area where the cis-epoxy ring of EPO-A is located.

Dihedral angle (°)	EPO-A*	EPO-tc
C1C2C3C4	-171.7	-169.5
C2C3C4C5	-58.6	-59.3
C3C4C5C6	-74.6	-72.5
C4C5C6C7	147.1	145.5
C5C6C7C8	-61.4	-60.1
C6C7C8C9	-68.2	-65.1
C7C8C9C10	168.1	171.4
C8C9C10C11	177.1	165.8
C9C10C11C12	***170.9***	***70.4***
C10C11C12C13	***-105.7***	***101.7***
C11C12C13C14	***-2.8***	***-145.6***
C12C13C14C15	98.0	93.4
C13C14C15O1	-75.4	-70.9
C14C15O1C1	149.7	134.0

Table 3. The torsion angles of the bound conformation of EPO-A and EPO-tc. *Reported values by Canales et al.[46]

Thus, as a further step, the obtained NMR data, including the derived epitope and the bound conformation, were employed to build a 3D model of bound EPO-tc. The obtained lowest energy conformer of EPO-tc is in agreement with the NMR experimental data was then docked into the described luminal binding site previously obtained for the EPO-A complex to generate the EPO-tc-tubulin complex (figure 17).

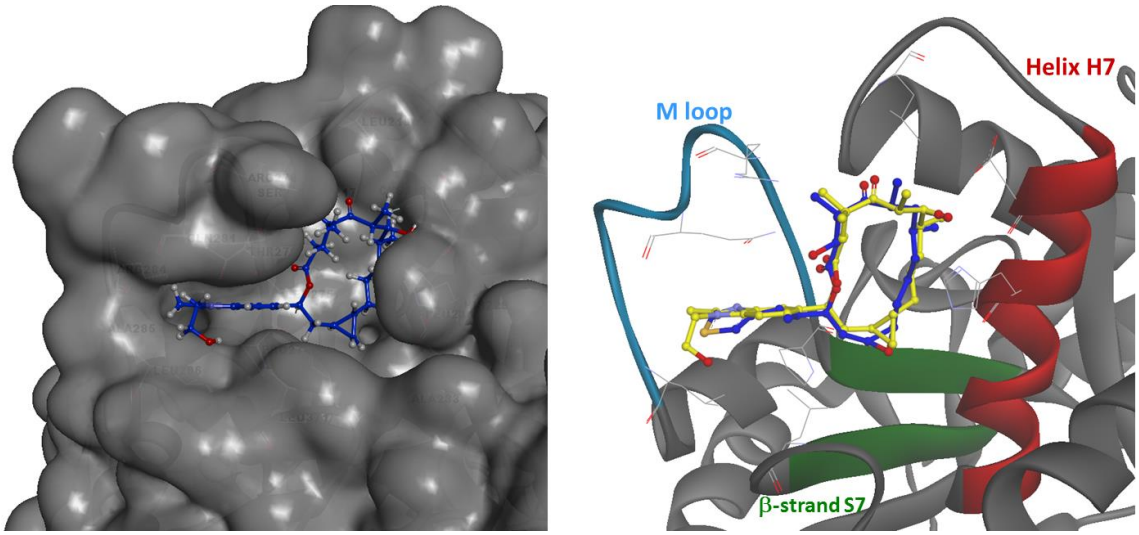


Figure 17. The 3D model for the EPO-tc:tubulin complex at the luminal binding site. Left panel: EPO-tc:tubulin complex. Right panel: superimposition of EPO-A (blue) and EPO-tc (yellow) at the luminal binding site.

The docking calculations were performed with AutoDock[47] and provided a similar binding mode to that described for EPO-A.[31,48] The binding pocket is described by the helix H7; β strand S7 and the M loop of β-tubulin subunit. EPO-tc establishes hydrogen bonds with His229, Asp226 and Thr276 while the rest of the binding pocket is predominantly hydrophobic. Therefore, the

presence of the trans-cyclopropane moiety in the epothilone derivative does not induce significant differences in the binding mode except local conformational adjustments around the trans-cyclopropane ring.

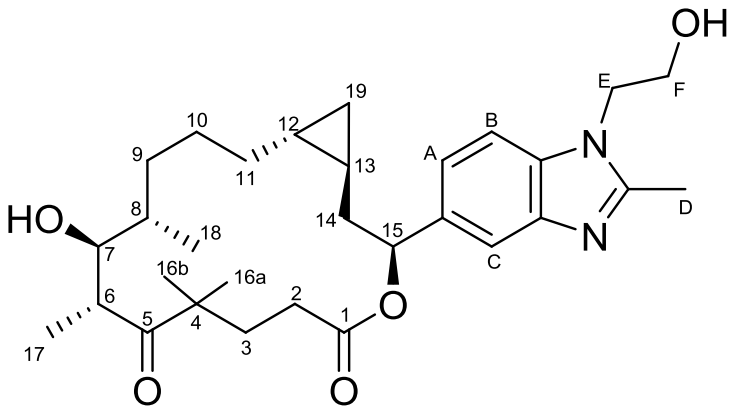
4.3.2. Conclusions

The combined NMR and molecular-modelling approach has permitted deducing that the EPO-tc derivative binds tubulin both in its dimeric and microtubule state of aggregation. The STD-derived epitopes were similar. The obtained bound conformations in both states were similar to those described for EPO-A with no significant changes in the binding modes.

4.3.3. Experimental

EPO-tc was provided by Prof. K-H Altmann from the Swiss Federal Institute of Technology (ETH) in Zürich, Switzerland.

The ^{1}H NMR resonances of EPO-tc (table 4) were assigned using standard NMR experiments (1D ^{1}H NMR, 2D TOCSY at different mixing times, 2D NOESY) in D_2O in a 600 MHz Bruker Avance spectrometer, equipped with a cryoprobe at 298 K and 310 K.



	^{1}H (δ, ppm)
H2a	1.89
H2b	2.38
H3a	1.96
H3b	2.08
H6	3.39
H7	3.67
H8	1.11
H9a	1.51
H9b	1.61

H10a	1.04
H10b	1.19
H11a	0.94
H11b	1.37
H12	0.77
H13	0.56
H14a	1.45
H14b	2.19
H15	5.98
H16a	0.93
H16b	1.34
H17	1.11
H18	0.90
H19	0.14
A	7.46
B	7.54
C	7.27
D	2.55
E	3.87
F	4.30

Table 4. The ^{1}H NMR resonances (δ, ppm) for EPO-tc at 298 K.

STD-NMR experiments were performed using a PBS (pH 7.3) deuterated buffered solution. A sample containing 10 μM of dimeric tubulin and 450 μM of EPO-tc (molar ratio 1:45) was prepared. The same molar ratio tubulin:EPO-tc (molar ratio 1:45) was used for the microtubules (10 μM) preparation. The microtubule sample was complemented with 0.1 mM GMPCPP6 as stabilizer. The STD experiments were acquired at 298 K for dimeric tubulin, and 310 K for microtubules, on a Bruker Avance 600 MHz spectrometer equipped with a cryoprobe. The saturation frequency was set at δ -1.5 ppm (aliphatic region) and the saturation time was 2 seconds.

Docking calculations were performed with AutoDock 4.2[47] with a 64 points box centered in the luminal binding site of β-subunit spacing 0.35 Å using a genetic algorithm. The coordinates of the crystal structure of the β-subunit (PDB[49] ID 4I50) were employed for the calculation while the ligand was built using Maestro from Schrodinger suite software.

4.4. The combined use of Dynamic Combinatorial Chemistry and NMR for searching new ligands

The search of novel ligands towards proteins has traditionally involved the chemical synthesis of compounds through reactions that result in the irreversible formation of new covalent bonds. However, recent interest has emerged on the employment of reversible reactions, under equilibrium control. Indeed, the application of dynamic combinatorial chemistry (DCC), which is based on a pool of reversible reactions, has become an interesting key tool in the drug discovery process.[50,51]

DCC is based on the knowledge of the behavior of a set of well-defined reversible reactions (dynamic combinatorial library, DCL). As a result of the reversibility, the reactants interconvert under thermodynamic control and are sensitive to the presence of external stimuli. In this context, the existence of interactions of the species present in a DCL-based mixture with a protein target should lead to the amplification of the formation of the interacting species (figure 18) according to Le Chatelier's principle. Following this concept, the drug discovery process may take the advantage of the template effect of therapeutic targets (i.e. proteins) on the particular DCL mixture giving rise to the synthesis of specific binders.[52–56]

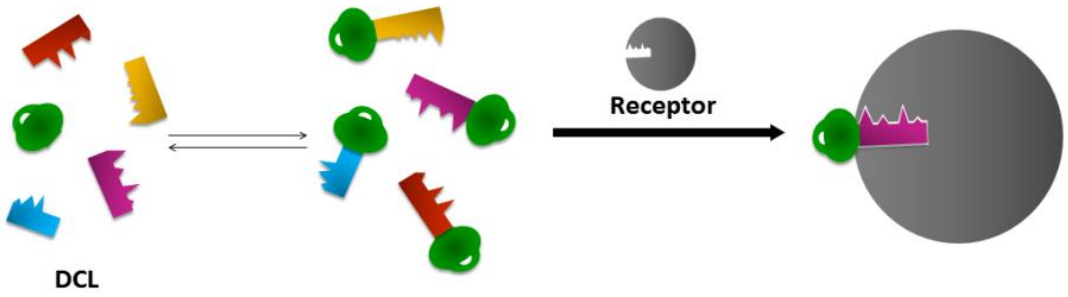


Figure 18. Scheme of the basis of DCC. DCL species reversibly react to form the products. In the presence of a target receptor, the formation of the particular molecule that selectively binds to the receptor is highly favored. The equilibrium is shifted and the synthesis of the bound molecule is preferred.

Thus, the choice of a suitable DCL is critical for the success of the DCC approach. Therefore, several reactions suitable for DCC have already been described.[52,57,58] Of course, there are a number of requisites that should be considered, given the presence of a protein template: pH, aqueous media, temperature, etc. Conversely, the reversible reactions of the DCL components should equilibrate fast enough under these conditions. The formation of acylhydrazones under

these conditions has already been described.[59–65] These molecules are suitable DCL components for protein templated DCC although the reaction process may be rather slow and require a nucleophilic catalyst for fostering the appropriate equilibration.[65]

There are relatively few contributions that have described the application of NMR techniques for in-situ monitoring of the success and outcomes of DCC methodology. As described throughout this Thesis, STD-NMR may provide the direct identification of the bound ligands to a given receptor, even employing a relatively complex mixture of molecules.[66] Moreover, being a ligand-based NMR method, STD-NMR requires a relatively small amount of protein (in the low micromolar range), while a large excess of ligand (in the high micromolar range, even milimolar) is usually required. Furthermore, additional ligand-based NMR experiments, such as tr-NOESY or DOSY, may also be employed in this context. Tr-NOESY experiments provide a direct proof of binding with the simple observation of the change in the sign of the NOESY cross-peaks of the ligand, passing from the free to the bound states. The analysis of these cross-peaks can be used to deduce the bound ligand conformation. Alternatively, DOSY monitors the changes in the translational diffusion coefficient upon complex formation.

In this section, as a proof-of-concept and to complement the different methodologies employed through this Thesis with a different perspective, ligand-based NMR experiments have been employed for the analysis of a DCL in the presence of a target protein. The DCL contained five different acylhydrazides and one aldehyde and could obviously produce the formation of 5 different acylhydrazones (figure 19). The chosen target protein was Frequenin-2 (Frq2), which is involved in synaptic connections and it is a therapeutic target for neurodegenerative diseases. This work has been performed in collaboration with Dr. Ruth Pérez at CIB-CSIC.

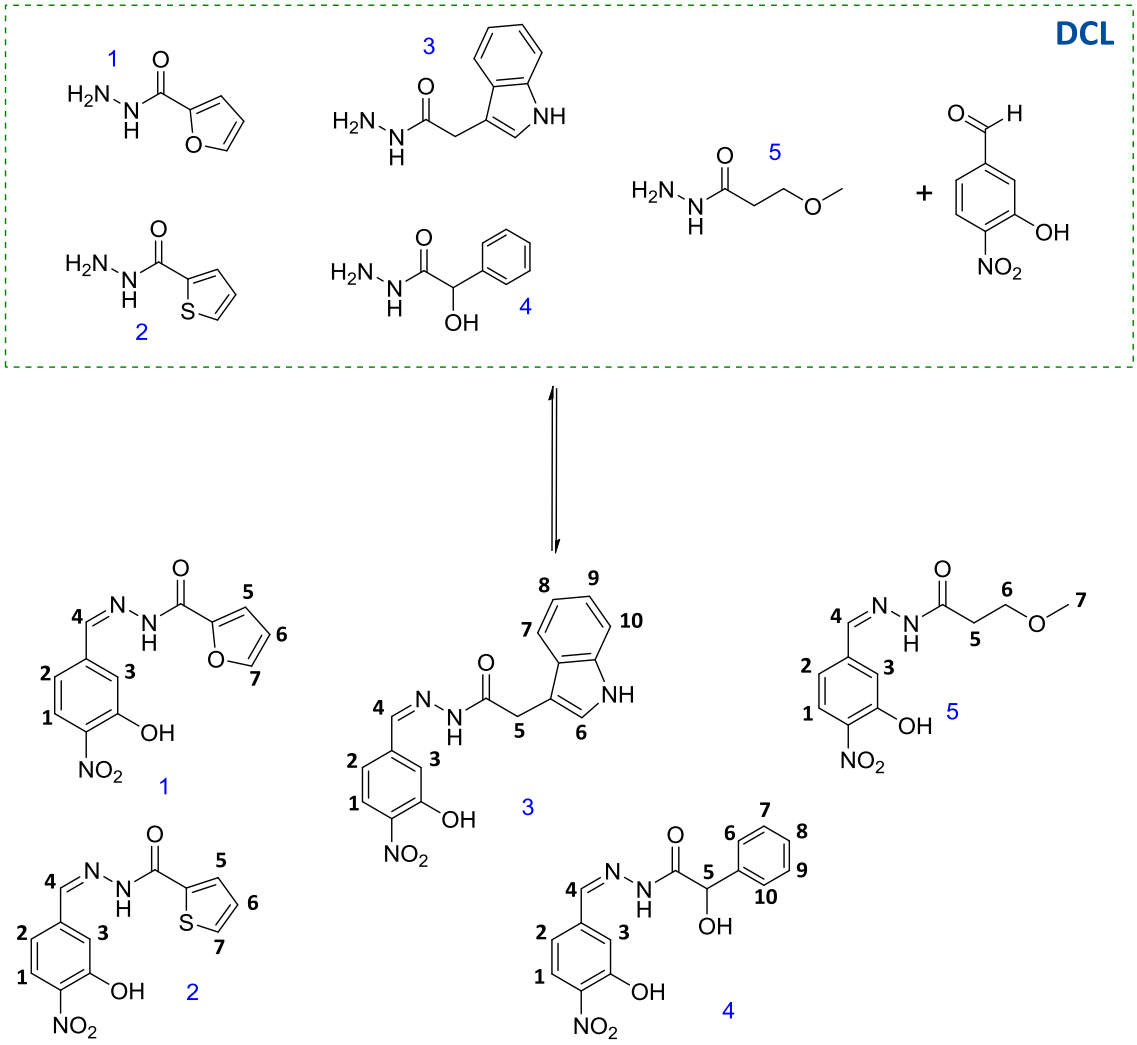


Figure 19. Scheme of the DCL mixture of acylhydrazides and the aldehyde together with the putative mixture of acylhydrazones.

4.4.1. Results and discussion

An initial set of experiments demonstrated that the reaction for the formation of the acylhydrazones does not equilibrate fast enough to accomplish the proper DCC conditions. Therefore, a nucleophilic catalyst (anisidine) was also added to the reaction mixture. A previous validation of the conditions was assessed by the acquisition of STD-NMR spectra of each product in the presence of the protein Frq2 (figures 20-24). In these samples, no catalyst was added to avoid the interconversion into the building blocks, aldehyde and corresponding acylhydrazide. The STD-NMR spectra revealed that all possible products except 5 were binders of the Frq2.

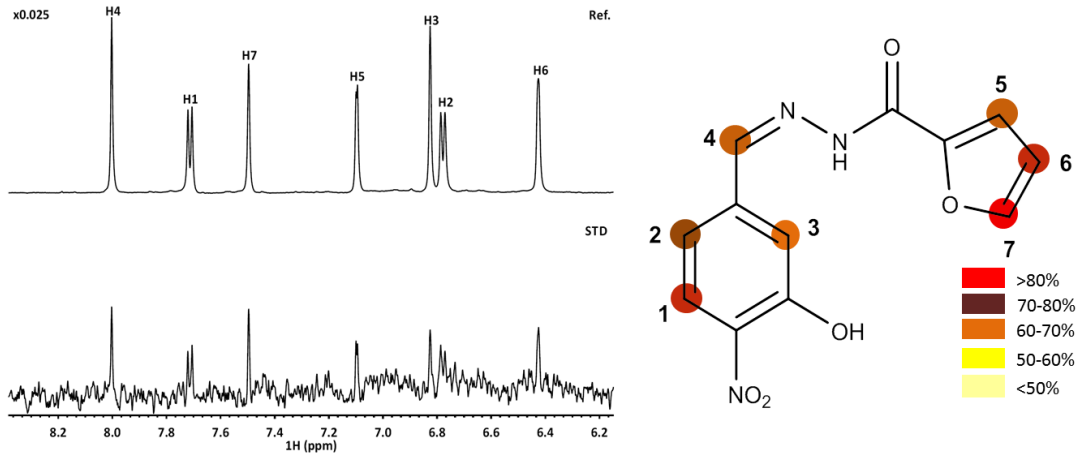


Figure 20. STD (below) and off-resonance (on top) NMR spectra of a 1mM sample of 1 in the presence of 10 µM Frq2 (281 K, 600 MHz).

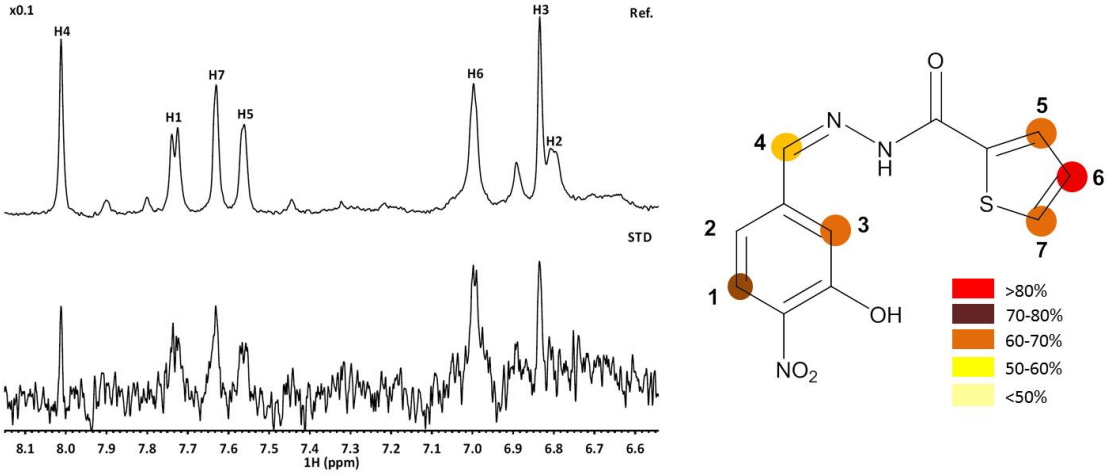


Figure 21.STD (below) and off-resonance (on top) NMR spectra of a 1mM sample of 2 in the presence of 10 µM Frq2 (281 K, 600 MHz).

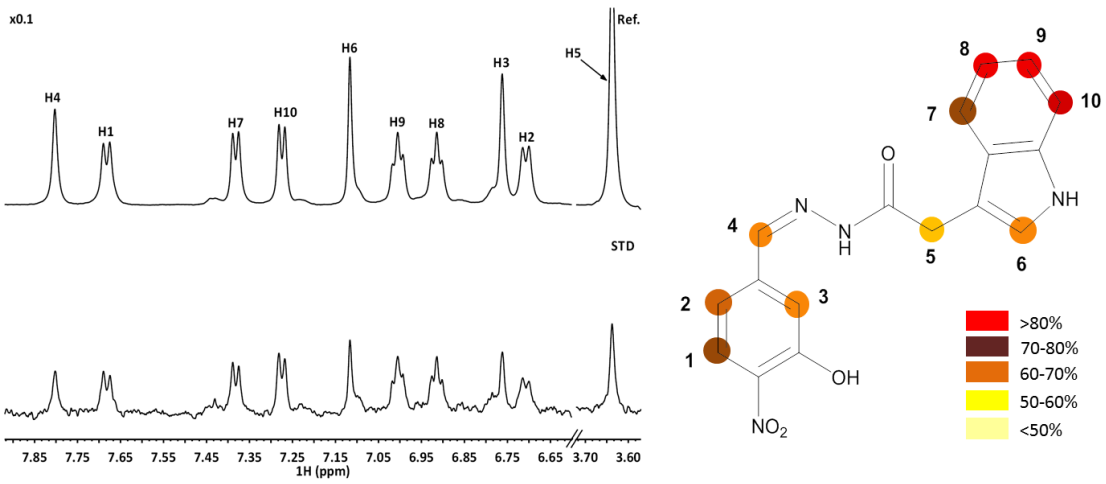


Figure 22. STD (below) and off-resonance (on top) NMR spectra of a 1mM sample of 3 in the presence of 10 µM Frq2 (281 K, 600 MHz).

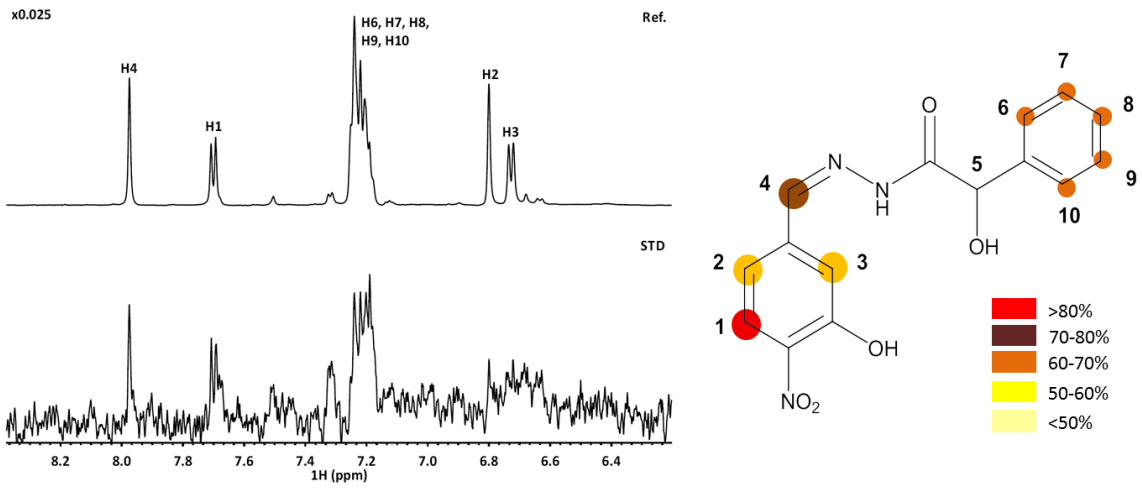


Figure 23. STD (below) and off-resonance (on top) NMR spectra of a 1mM sample of 4 in the presence of 10 μM Frq2 (281 K, 600 MHz).

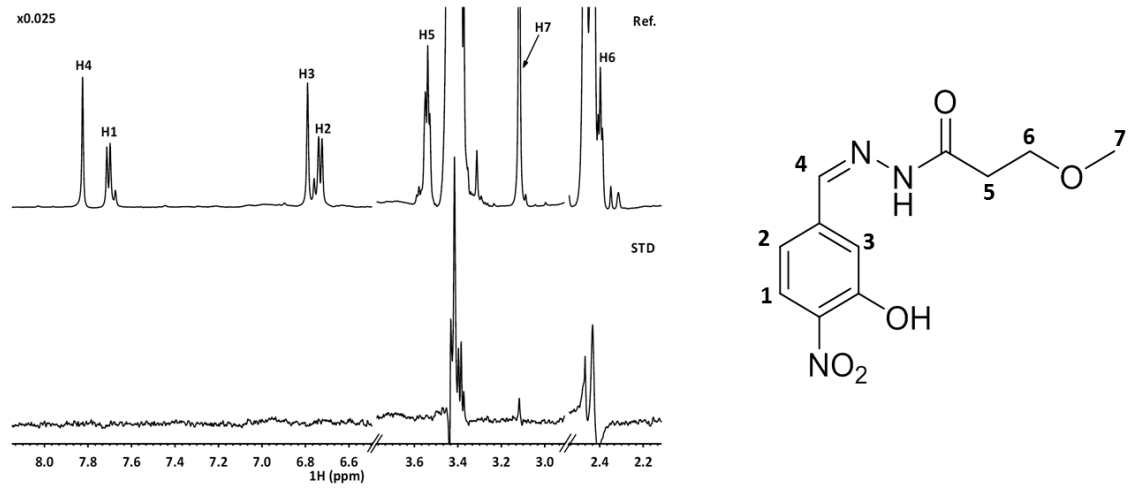


Figure 24. STD (below) and off-resonance (on top) NMR spectra of a 1 mM sample of 5 in the presence of 10 μM Frq2 (281 K, 600 MHz).

The analysis of the STD-NMR spectra revealed that **3** displays the largest absolute STD intensities (best binder) while **5** is not recognized by Frq2 under the employed conditions. In principle, it should be expected that the formation of **3** would be favored under DCC conditions in the presence of Frq2. Therefore, the DCL mixture was equilibrated in the presence of Frq2 and analyzed by NMR methods.

The observation of the simple ^{1}H NMR (figure 25) spectrum of the DCL mixture after equilibration showed the formation of the different products. Indeed, the analysis of the DOSY spectra revealed that the formed products displayed large diffusion coefficient (figure 25) in the presence of Frq2, demonstrating the formation of ligand:protein complexes. An additional verification of the existence of a molecular recognition process was achieved by performing tr-

NOESY experiments. In this case, negative NOE cross-peaks were observed for the products of the mixture, while the reactants and catalyst displayed positive and quantum zero NOE cross peaks (figure 27).

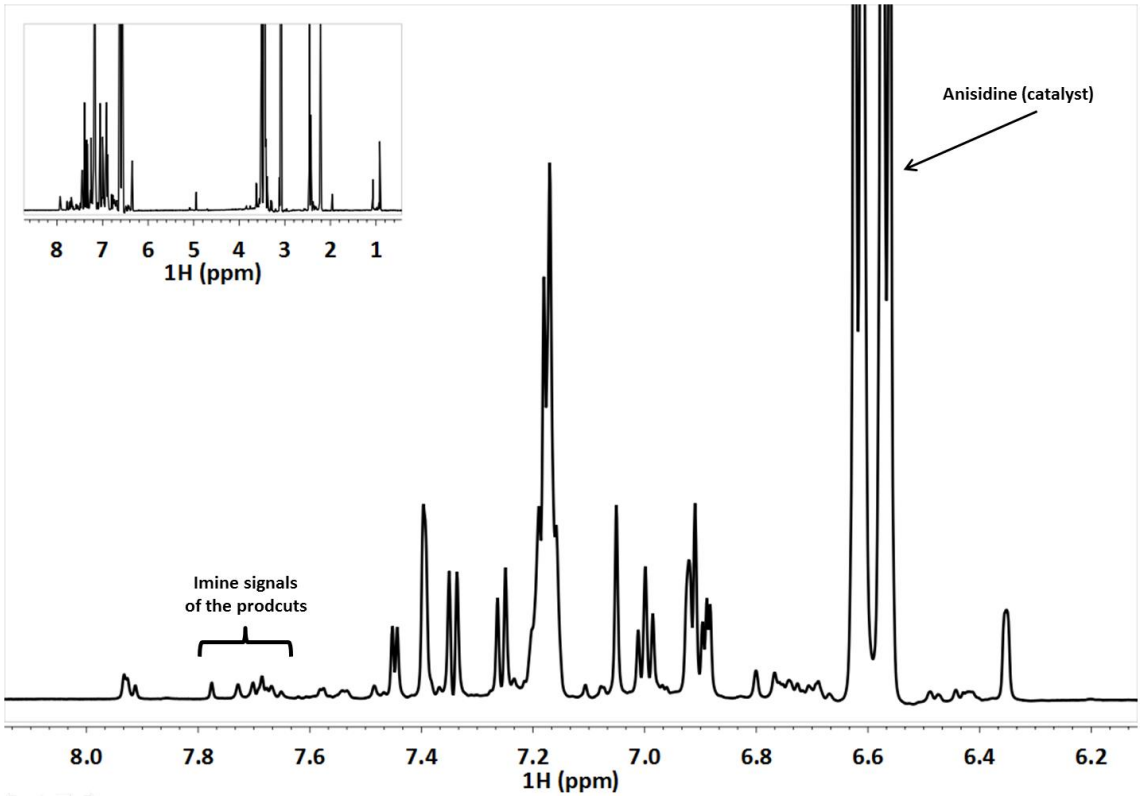


Figure 25. 1D ^{1}H NMR spectrum of the mixture in the presence of Frq2 (281 K, 600 MHz). Weak signals arising from several formed products were observed.

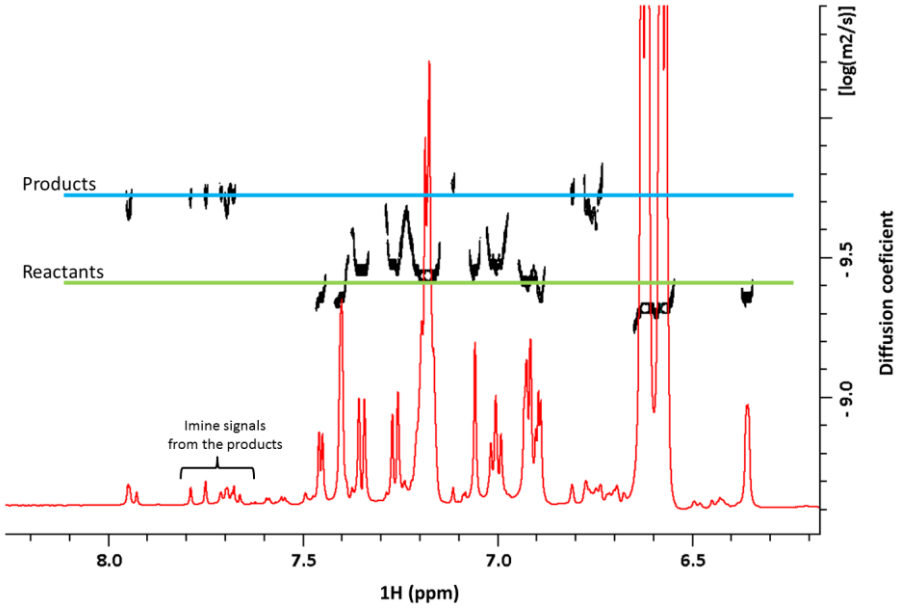


Figure 26. DOSY experiment of the DCL in the presence of Frq2 (black) and ^{1}H NMR spectrum (red) (281 K, 600 MHz). The signals from the products display significantly larger diffusion coefficient than those of the reactants due to formation of protein:ligand complex

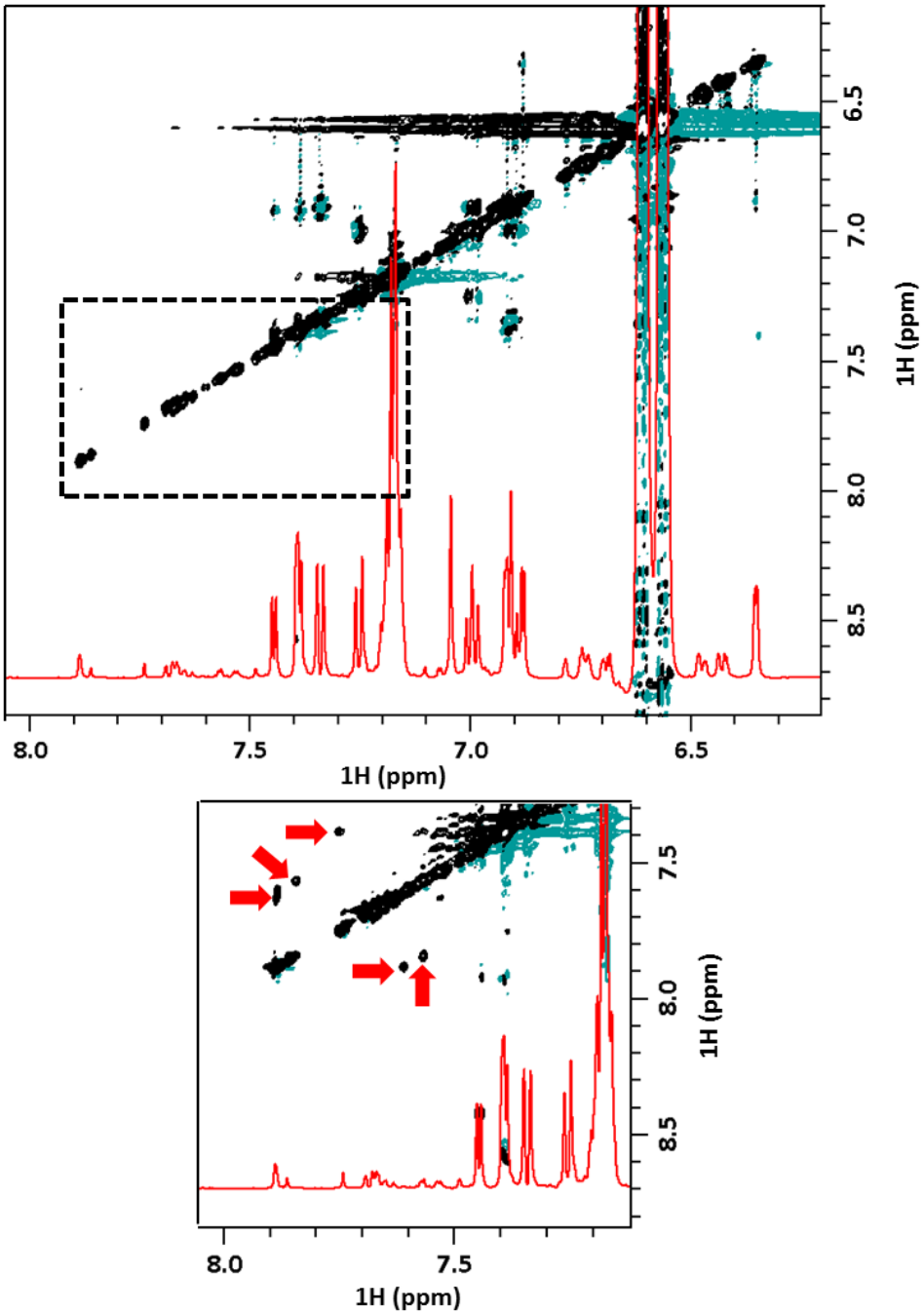


Figure 27. tr-NOESY spectrum of DCL mixture in the presence of Frq2 with a mixing time of 200 ms (281 K, 600 MHz). At the bottom: Amplification of the region δ 7.2-8 ppm. The negative NOE cross-peaks corresponding to intermolecular NOEs of the products are highlighted with red solid arrows.

Nevertheless, the observed signals corresponded to several products. No clear evidence was found for the formation of a unique product from this initial mixture. Thus, a simplified DCL was studied in the presence of Frq2.

The chosen simplified DCL consisted on a mixture of product **5** and the starting building block of the best binder, **3**. In this case, the existence of the equilibrium process of **5** would release aldehyde, which could concurrently be used for the formation of product **3**. In the presence of

Frq2, only the equilibrium of the formation of **3** would be affected, driving the DCL towards the formation of **3**.

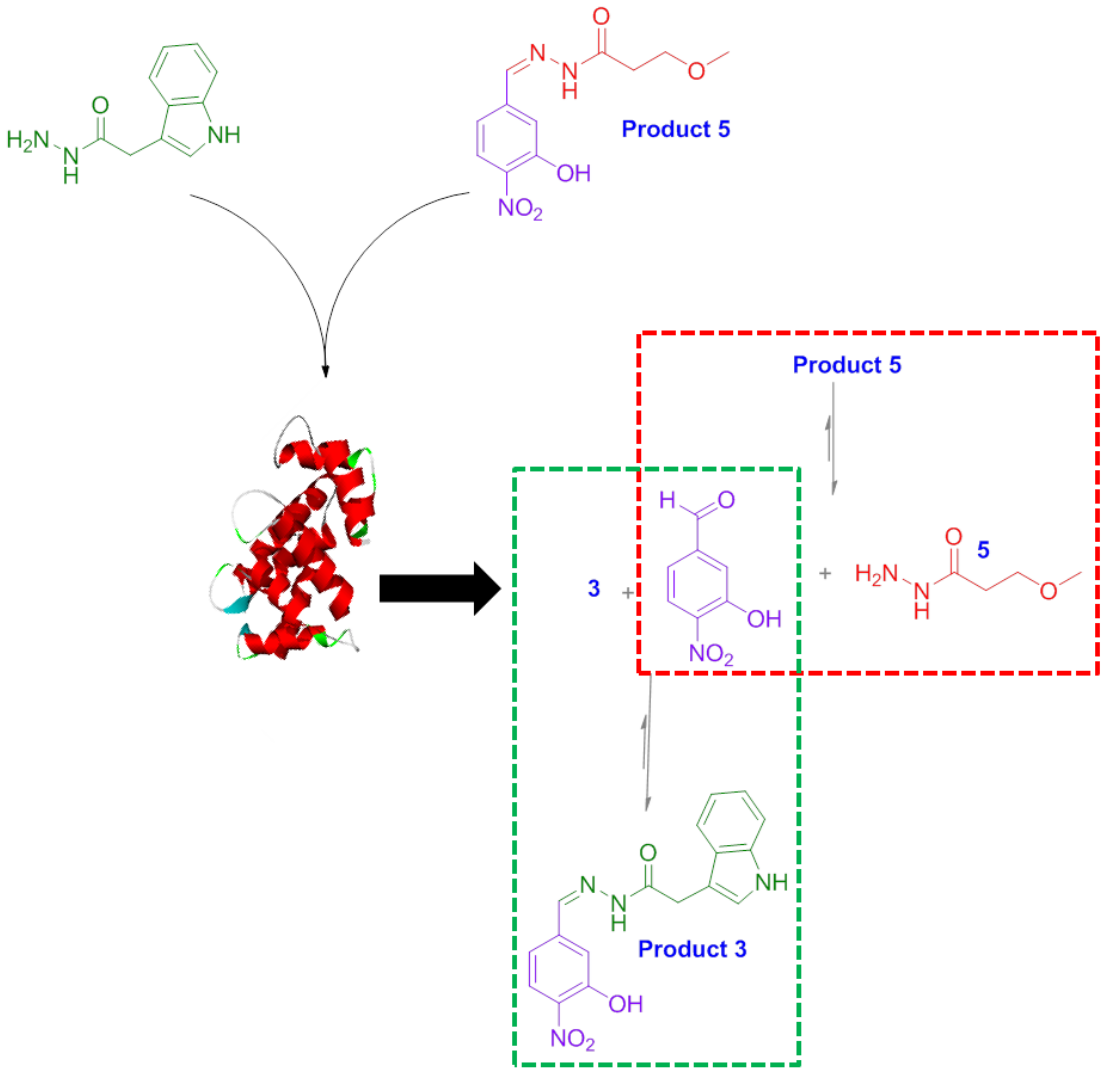


Figure 28. Scheme of the simplified DCL.

Indeed, the analysis of the ^{1}H NMR spectra at t=0 and after 15 days revealed the presence of new weak signals at the aromatic region assigned to **3** and a new triplet at δ 2.2 ppm that belongs to H6 of the acylhydrazide precursor of **5**. These new signals evidenced the progression of the mixture towards the formation of a unique product, **3**, due to the amplification produced by Frq2.

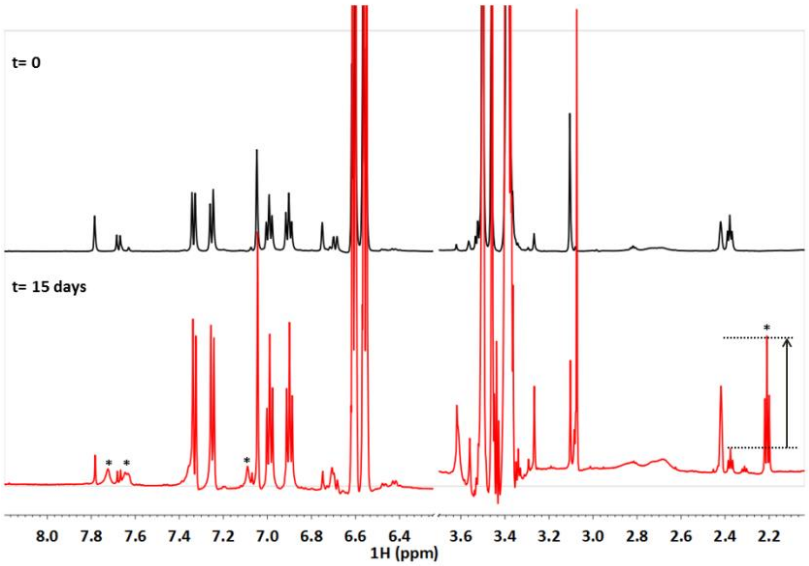


Figure 29. ^{1}H NMR spectrum of the sample 3+5+Frq2 at different reaction times (up to 15 days) (281 K, 600 MHz).The new signals that reveal the formation of 3 (aromatic region) and 5 (aliphatic region) are marked with stars.

In fact, the analysis of the NMR spectra after 15 days showed that only the weak signals corresponding to **3** display STD effects, validating this DCL system.

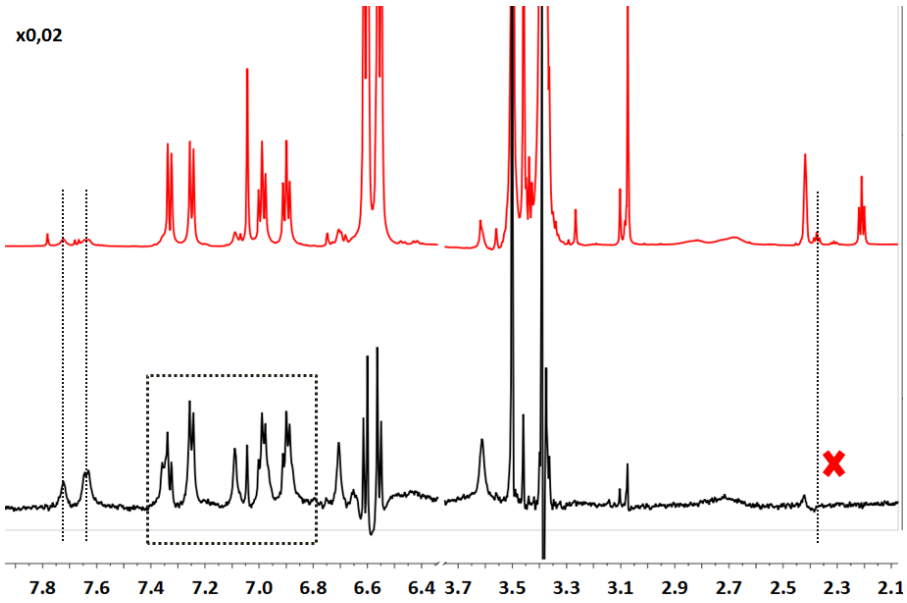


Figure 30. STD (below) and off-resonance (on top) NMR spectra of the mixture 3+5+Frq2 (281 K, 600 MHz) after 15 days.

Finally, the binding site of **3** was assessed by using STD-NMR-based competition studies versus chlorpromazine (CPZ), a well-known Frq2 binder (PDB[49] ID 5G08) (figure 31). When increasing amounts of CPZ were added, the STD intensities signals of **3** significantly decreased, strongly suggesting that **3** and CPZ compete for the same protein binding pocket.

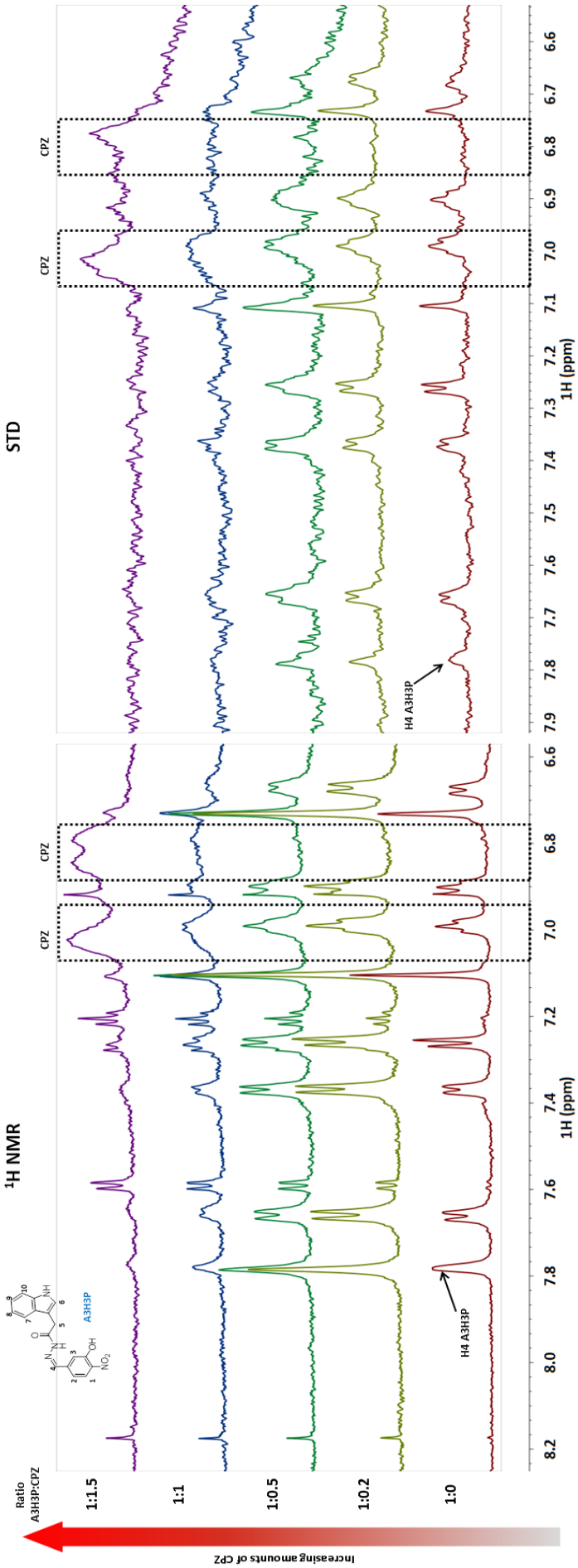


Figure 31. The competition of 3 versus CPZ in the presence of Frq2, as deduced from STD-NMR based competition experiments. The ratio of CPZ is increased from bottom to top.

4.4.2. Conclusions

The employed DCL in the presence of Frq2 provided STD signals, except for **5**, strongly suggesting that the presence of aromatic rings has a key role for Frq2 binding. The analysis of the DOSY experiments permitted to assess that the formed products interact with the target protein, since they display much larger diffusion coefficient than the initial reactants. This fact can be explained by the formation of protein:ligand complexes. Furthermore, tr-NOESY experiments also validated the existence of binding.

The existence of amplification of one particular product from the DCL system was elusive when using the complete library. Thus, a simplified DCL was employed to reveal that product **3** is bound and amplified by Frq2 in the presence of **5**. This fact validates the role of Frq2 as a selective template. Finally, STD-NMR-based competition studies demonstrated that the amplified product **3** competes for the same binding site than a well-known Frq2-binder, CPZ.

4.4.3. Experimental

STD-NMR experiments were performed using a perdeuterated TRIS (pH 7.9) buffered solution with an aliquot of d6-DMSO. A sample containing 10 μM of Frq2 and 1 mM of DCL species (molar ratio 1:100) was prepared. The STD experiments were acquired at 281 K on a Bruker Avance 600 MHz spectrometer equipped with a cryoprobe. The saturation frequency was set at δ -0.5 ppm (aliphatic region) and the saturation time was 2 seconds. A spin lock filter was applied to minimize signals from Frq2. The same conditions were used for the acquisition of the STD-NMR spectra of the individual products with Frq2.

tr-NOESY experiment and DOSY experiments were performed on the same samples at 281 K on the same spectrometer. The tr-NOESY mixing time was fixed at 200 ms. The DOSY experiments were acquired with 16 gradient increments to a final intensity decay of 90%.

- STD-based competition studies

To a solution of 10 μM of Frq2 and 300 μM of **3**, increasing amounts of CPZ were added, until a final concentration of 450 μM of CPZ was reached (table 1). The STD-NMR spectra were recorded in a 600 MHz spectrometer equipped with a cryoprobe at 281 K. The saturation frequency was set at δ -0.5 ppm (aliphatic region) and the saturation time was 2 seconds.

	Frq2 (μM)	3 (μM)	CPZ (μM)
Exp. 1	10	300	0
Exp. 2	10	300	60
Exp. 3	10	300	150
Exp. 4	10	300	300
Exp. 5	10	300	450

Table 5. Description and summary of the experiments and concentrations employed for the STD-NMR-based competition assays with Frq2.

4.5. NMR & ligand-DNA interactions

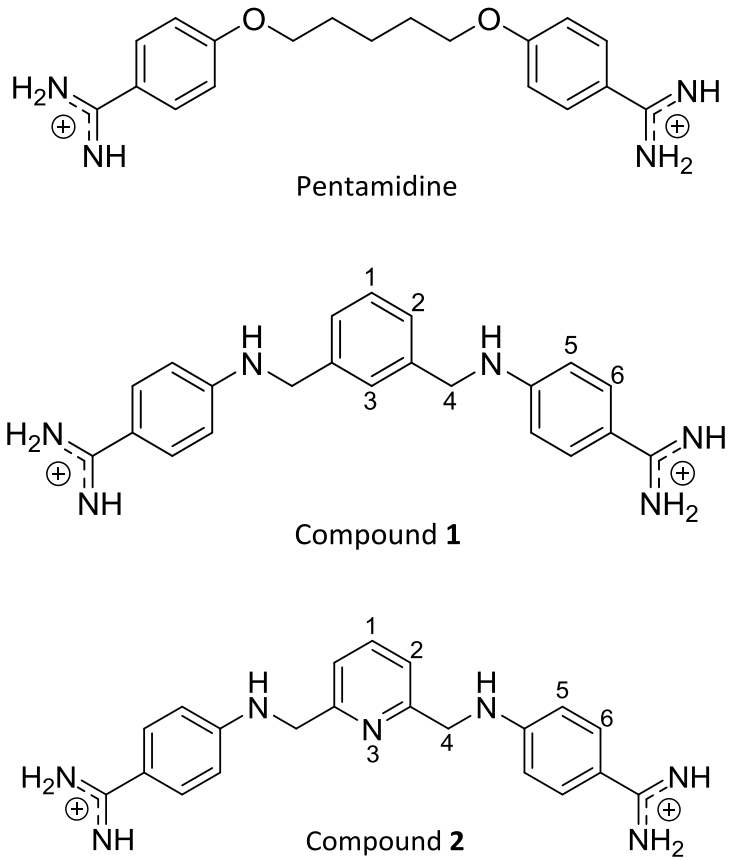
Transcription factors (TFs) translate the information encoded in the DNA. Indeed, they play a key role in the regulation of gene expression[67] and its alteration may triggers severe genetic pathologies such as cancer.[68–71] Therefore, the development of sequence-selective synthetic molecules is of interest for advances in genetic medicine.[72] In this context, the rational design of TF-mimics synthetic peptides that are selective towards specific DNA sequences has to overcome two major challenges: DNA affinity and selectivity. TFs display extremely selective binding to a consensus DNA sequence out of the vast of alternative sites that also display high affinity.[73] At this point, one of the most useful approach for synthetic DNA binding peptides involve the use of small molecules, selective DNA binders, which may be conjugated with a given peptide domain. In fact, different small molecules have been described to target different DNA sites: intercalators bind DNA between base pairs,[73,74] major groove binders bind at the major groove of the duplex helix,[75,76] while minor groove binders have also been developed.[77,78] The essence of the interaction of these small molecules with DNA resides in a complementary network of intermolecular hydrogen bonds, hydrophobic interactions and electrostatic forces together with the appropriate geometry of the ligand, suitable for accommodation at the DNA site.[79,80]

The DNA minor groove displays different electrostatic potentials, depending on the nucleotide sequence. The minor groove wall of A/T-rich sequences shows negative potentials while, G/C-rich sequences display positive potentials. In addition, A/T base pairs offer different number of possible hydrogen bonds than G/C. These topologic considerations have led the design of A/T- or G/C-selective molecules. In fact, the A/T-selective molecules at the minor groove share common features such as positive charges, number of hydrogen donors and acceptors and a concave shape.[80]

In this section, we have investigated some features of DNA minor groove binders, designed and synthesized by the group of Prof. J.L. Mascareñas and E. Vázquez-Sentís at CIQUS

(University of Santiago de Compostela), by using NMR methods in combination with computational tools. The DNA sequence used for this study is a well-known model that consists in a double stranded (ds) dodecamer of base pairs with a particular palindromic sequence, d(CGCGAATTCGCG), the Dickerson's dodecamer.[81]

The two employed molecules (**1** and **2**, scheme 1) are inspired in the antimicrobial pentamidine, a well-known minor groove binder, and are characterized by possessing a suitable concave shape for minor groove binding, while the two amidinium moieties are positively charged providing stabilizing electrostatic interactions with A/T-rich sequences.[82] They differ in the chemical nature of the central aromatic moiety (phenyl or pyridine).



Scheme 1. Scheme of the chemical structures of the synthetic compounds 1, 2 and the pentamidine.

In this case, due to the size of the DNA-ligand complexes, it was possible to apply NMR experiments based in the observation of both, the ligand and the DNA receptor, by tr-NOESY and their intermolecular interactions in order to obtain structural information.

4.5.1. Results and discussion

As a first step the ^{1}H NMR resonances of the Dickerson's dodecamer were assigned under our experimental conditions (see experimental section) and were in accordance with previous reported chemical shifts.[83,84] Next, a chemical shift perturbation (CSP) strategy was applied. For any ligand or receptor molecule, the presence of its interacting partner provides perturbations of the chemical environment of the NMR-active nuclei, triggering changes in the ^{1}H NMR chemical shifts. Fittingly, the largest changes in the chemical shifts of the receptor´s nuclei reveal the most affected ones due to the presence of the ligand. Thus, the interaction between the two compounds and the Dickerson dodecamer was analyzed by monitoring the CSP of the ^{1}H NMR signals of the DNA sequence upon addition of **1**-**2**. A series of NOESY experiments were employed to monitor the CSPs of the DNA signals and at the same time to monitor the possible intermolecular contacts (figures 32-35).

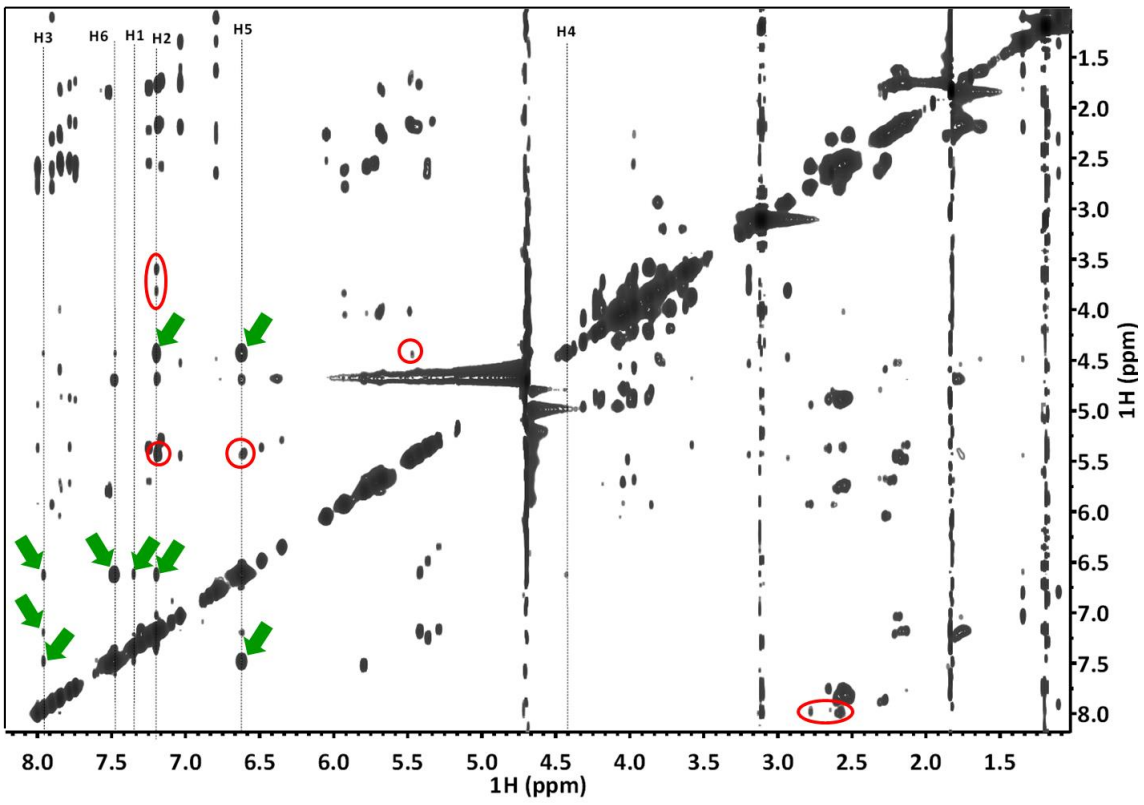


Figure 32. tr-NOESY of the DNA:1 complex (300 K, 500 MHz). Some ligand intramolecular NOE crosspeaks are annotated with green solid arrows. Some intermolecular NOE crosspeaks are annotated with red circles.

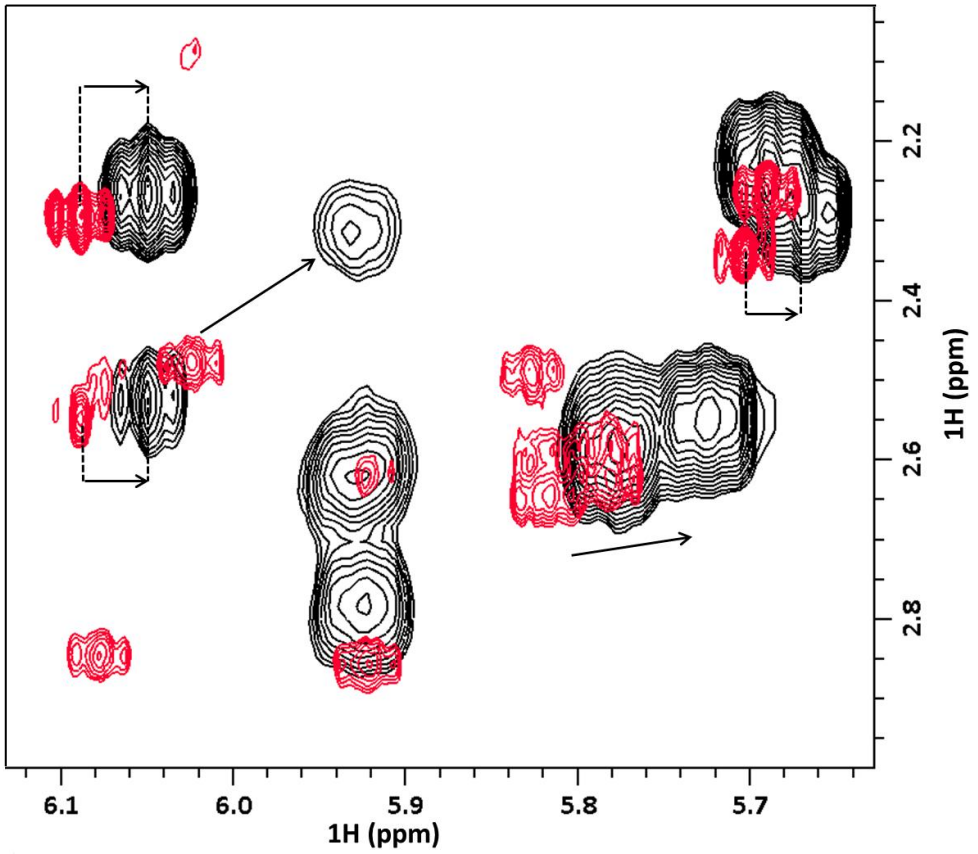


Figure 33. The interaction of 1 with DNA. Superimposition of the NOESY spectra recorded for DNA in the absence (red) and presence of 1 mM of compound 1 (black).

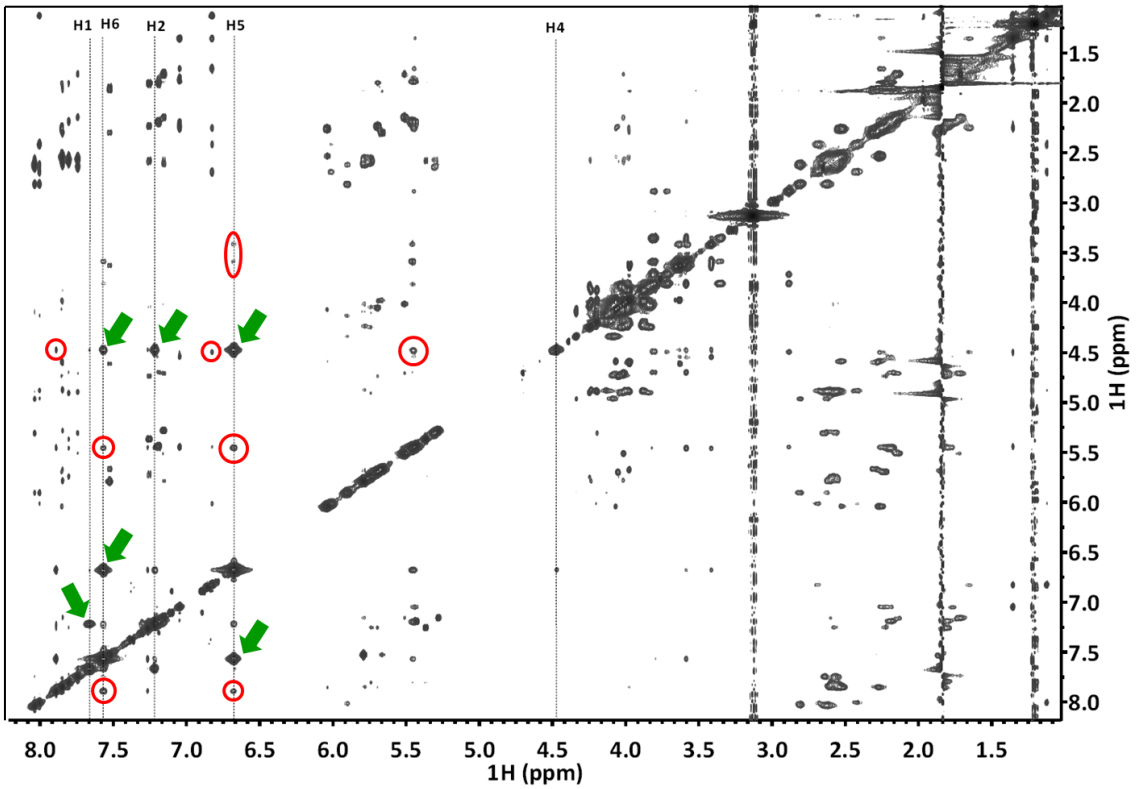


Figure 34. tr-NOESY of the DNA:2 complex (300 K, 500 MHz). Some intramolecular NOE crosspeaks are annotated with green solid arrows. Some intermolecular NOE crosspeaks are annotated with red circles.

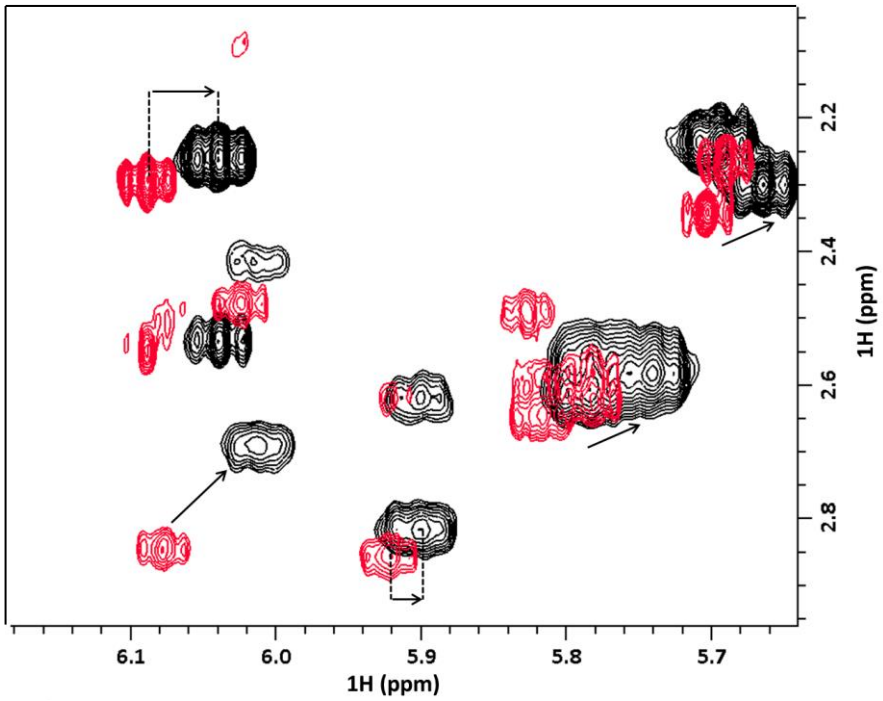


Figure 35. The interaction of 2 with DNA. Superimposition of the NOESY spectra recorded for DNA in the absence (red) and presence of 1 mM of compound 2 (black).

The CSP analysis of the DNA:ligand (1:1 molar ratio) complex revealed that the most affected protons in complexes with **1** (figure 36) and **2** (figure 37) belong to the central part of the dodecamer d(-$A_6T_7T_8C_9$-), rich in adenines and thymines. Indeed, protons displaying largest CSP were H4' of T_7, T_8, A_6 and C_9, oriented towards the DNA minor groove, while those exposed to the major groove did not display any significant perturbation, H3' of A_6, H7 of T_7, H5 of C_9 or H3' of T_8.

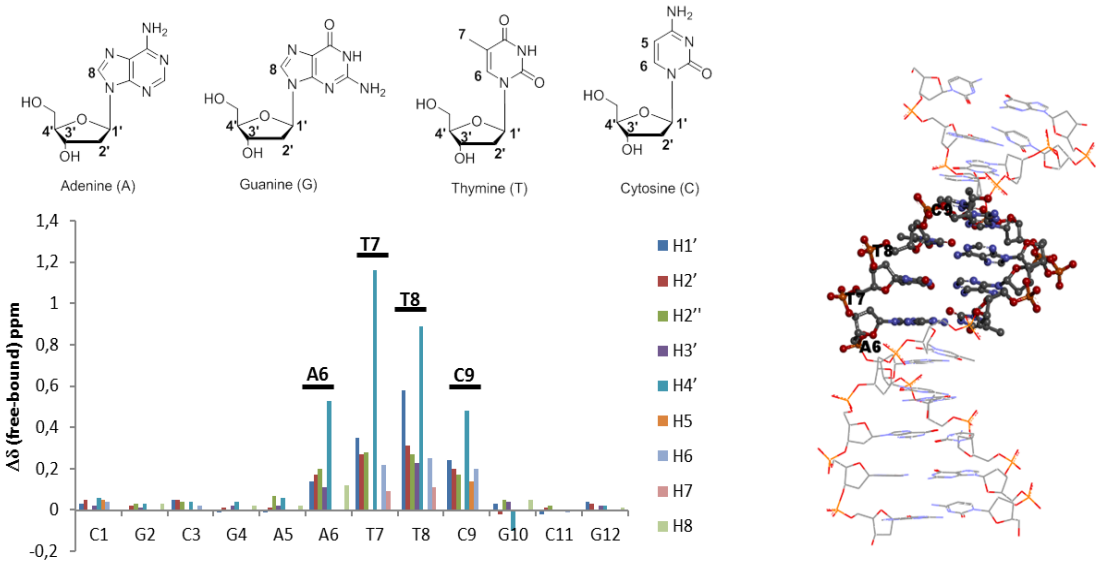


Figure 36. Chemical shift perturbation of ^{1}H NMR signals of the Dickerson´s dodecamer in the presence of 1.

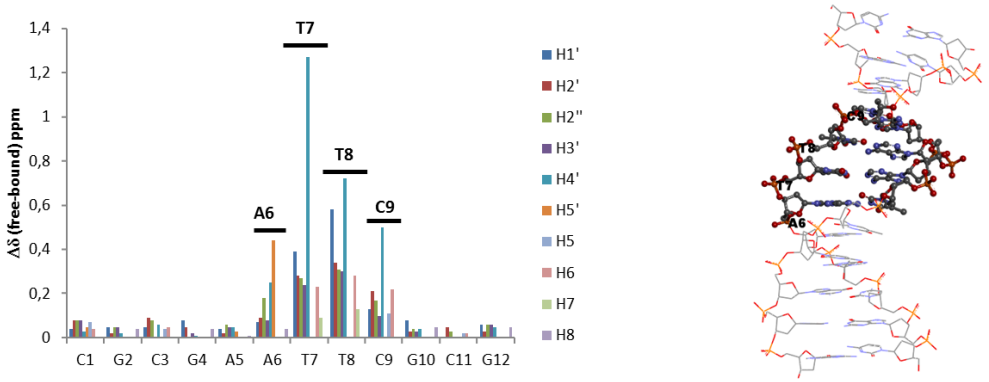


Figure 37. Chemical shift perturbation of ^{1}H NMR signals of the Dickerson´s dodecamer in the presence of 2.

Furthermore, a 3D model of both ligand-DNA complexes were then generated by docking calculations using AutoDock[47] (figures 38 and 39). The obtained structures were qualitatively validated with the information provided by the assigned intermolecular NOEs (tables 6 and 7).

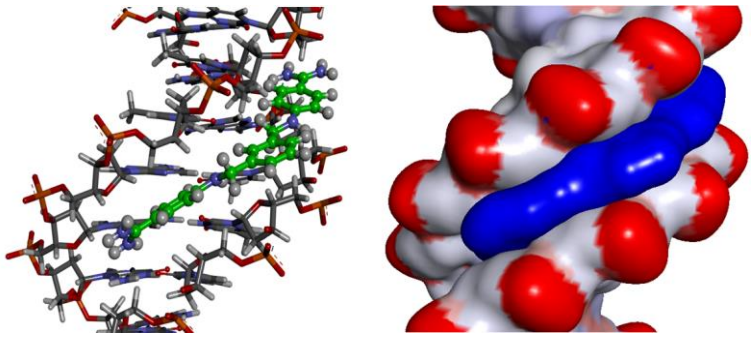


Figure 38. 3D model of the interaction of 1 with Dickerson´s dodecamer.

Compound 1	DNA proton	distance in model (Å)	
		(chain A 3´-5´)	(chain B 5´-3´)
H5	H1´ A6	3.5	-
	H1´ T7	3.5	3.3
	H5´ T7	3.5	-
	H4´ C9	3.3	-
	H4´ T8	-	3.6
H6	H5´ T6	3.2	3.0
	H4´C9	-	3.0
	H1´ C9	3.7	2.8
	H1´ A6	2.2	3.4
H4	H4´ T8	-	3.3
	H4´ T7	2.4	3.4
H1	H4´ C9	4.0	-

Table 6. List of observed intermolecular NOEs in the complex 1:DNA and corresponding distances in the model.

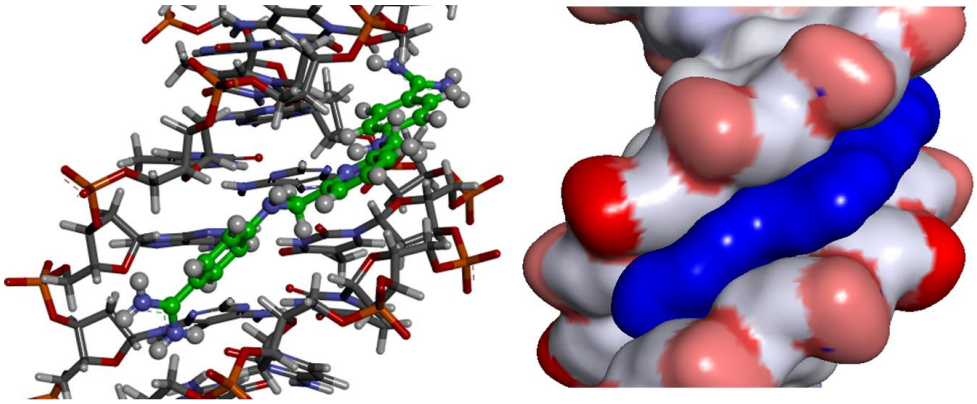


Figure 39. 3D model of the interaction of 2 with Dickerson´s dodecamer.

Compound 2	DNA proton	distance in model (Å)	
		(chain A 3´-5´)	(chain B 5´-3´)
H5	H2 A6	2.2	2.3
	H1´ T7	3.3	3.2
	H5´ T7	3.7	3.6
	H4´ C9	3.2	3.1
	H5'C9	3.8	3.4
H6	H5´ T7	3.3	3.2
	H4´C9	3.0	2.8
	H1´ C9	3.0	3.1
	H2 A6	2.2	2.3
H4	H4´ T8	-	3.3
	H4´ T7	2.4	3.4
H1	H4´ C9	4.0	-

Table 7. List of observed intermolecular NOE in the complex 2:DNA and corresponding distances in the model.

The analysis of the models revealed that, for both cases, the terminal amidinium moieties establish hydrogen bonds with the DNA and CH-pi hydrophobic interactions can also take place between ribose H4 and the aromatic rings. For **2**, the possibility of an additional hydrogen bond between the nitrogen of the pyridine and the nucleotides could not be further validated. No significant differences could be observed between both models. In fact, the affinity results obtained by collaborators at Santiago de Compostela revealed that **2** only displays a slightly higher affinity (K_D= 1.0 μM) than **1** (K_D= 1.8 μM).[85] This comparable affinity also supports a fairly similar binding mode for both compounds.

4.5.2. Conclusions

Ligands **1** and **2** display certain specificity towards A/T-rich regions of DNA establishing mainly hydrogen bond interactions with the base pairs at the minor groove of a dsDNA helix. The presence of cyclic nitrogen (**2**) does not provide significant changes in the binding mode.

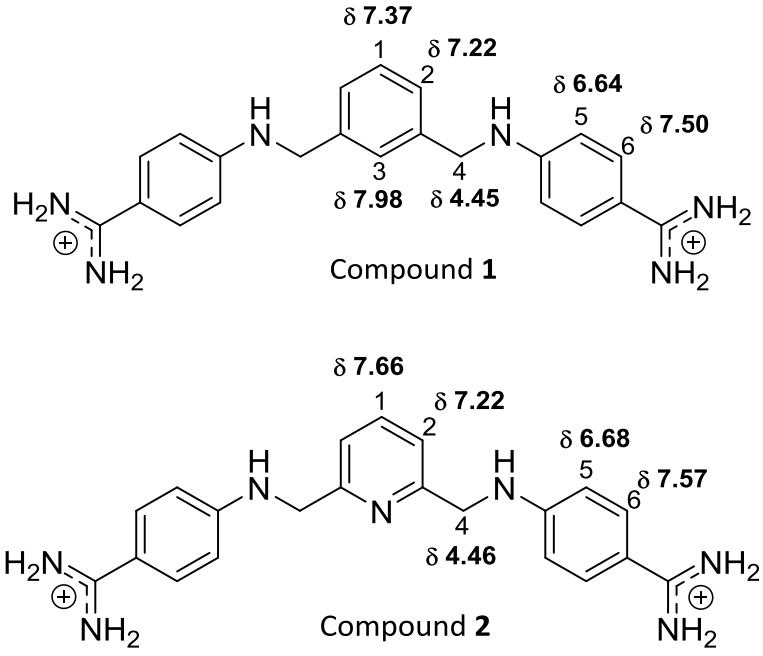
4.5.3. Experimental

Compounds **1** and **2** were synthesized by Prof. Mascareñas and Vázquez-Sentís group from CIQUS, Santiago de Compostela.

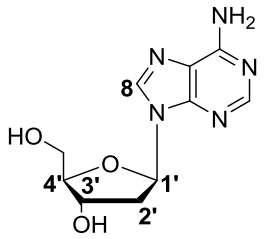
The Dickerson dodecamer ds(CGCGAATTCGCG) was purchased from Thermo-Scientific.

The ^{1}H NMR resonances of DNA dodecamer and ligand **1** and **2** were assigned by the acquisition of a set of standard NMR experiments (1D ^{1}H-NMR, 2D TOCSY at different mixing times, 2D NOESY) in D_2O in a Bruker Avance 500 MHz spectrometer at 300 K. For the assignments of DNA NHs resonances experiments were performed in H_2O as well.

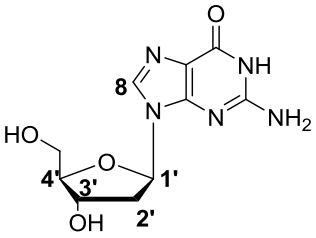
tr-NOESY experiments of the mixtures of 1.1 mM of **1** or **2** with DNA (1:1 molar ratio) were performed in H_2O and D_2O at 300 K and a mixing time of 100, 200 and 400 ms in a Bruker Avance 500 MHz spectrometer.



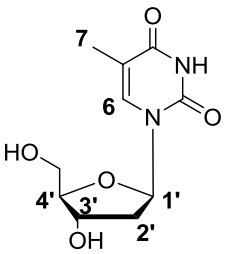
Scheme 2. Chemical representation of the structures of compound 1 to 2. ^{1}H chemical shift values are annotated.



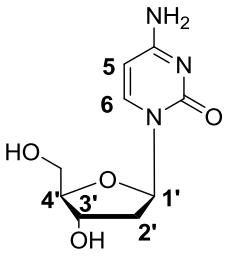
Adenine (Ade, A)



Guanine (Gua, G)



Thymine (Thy, T)



Cytosine (Cyt, C)

	^1H (δ, ppm)	H1'	H2'	H2''	H3'	H4'	H5	H6	H7	H8
Free DNA	*Cyt1*	5.69	1.89	2.32	4.62	4.02	5.84	7.56		
	Gua2	5.78	2.56	2.63	4.89	4.26				7.87
	Cyt3	5.53	1.78	2.19	-	4.05	5.29	7.18		
	Gua4	5.36	2.57	2.66	4.91	4.23				7.76
	Ade5	5.91	2.60	2.85	4.97	4.37				8.02
	Ade6	6.07	2.48	2.84	4.91	4.38				8.02
	Thy7	5.81	1.90	2.47	-	4.10		7.02	1.19	
	Thy8	6.01	2.08	2.46	4.78	4.09		7.29	1.45	
	Cyt9	5.57	1.97	2.32	-	4.06	5.55	7.38		
	Gua10	5.76	2.51	2.61	4.91	4.07				7.83
	Cyt11	5.67	1.81	2.24	-	4.05	5.36	7.24		
	Gua12	6.08	2.29	2.53	4.60	4.08				7.86
DNA:1 complex	*Cyt1*	5.66	1.84	2.89	4.60	3.96	5.79	7.52		
	Gua2	5.78	2.54	2.60	4.88	4.23				7.84
	Cyt3	5.48	1.73	2.15	-	4.01	5.29	7.16		
	Gua4	5.37	2.56	2.66	4.89	4.19				7.74
	Ade5	5.92	2.59	2.78	4.95	4.31				8.00
	Ade6	5.93	2.31	2.64	4.80	3.85				7.90
	Thy7	5.46	1.63	2.19	-	2.94		6.80	1.10	
	Thy8	5.43	1.77	2.19	4.55	3.20		7.04	1.34	
	Cyt9	5.33	1.77	2.15	-	3.58	5.41	7.18		
	Gua10	5.73	2.53	2.56	4.87	4.17				7.78
	Cyt11	5.69	1.80	2.22	-	4.3	5.36	7.25		
	Gua12	6.04	2.26	2.53	4.58	4.06				7.85

DNA:2 complex	*Cyt1*	5.66	1.80	2.24	4.57	3.97	5.79	7.53		
	Gua2	5.77	2.52	2.26	4.84	4.23				7.84
	Cyt3	5.49	1.68	2.12	4.72	4.00	5.26	7.14		
	Gua4	5.30	2.54	2.70	4.89	4.22				7.74
	Ade5	5.88	2.60	2.78	4.93	4.33				8.02
	Ade6	6.00	2.39	2.66	4.85	4.13				7.99
	Thy7	5.43	1.63	2.22	4.48	2.84		6.80	1.10	
	Thy8	5.44	1.75	2.16	4.51	3.39		7.02	1.33	
	Cyt9	5.45	1.78	2.15	4.70	3.57	5.44	7.17		
	Gua10	5.72	2.54	2.56	4.89	4.23				7.79
	Cyt11	5.69	1.78	2.22	-	4.06	5.35	7.24		
	Gua12	6.03	2.50	2.24	4.55	4.05				7.83

Table 8. ^{1}H NMR resonances of DNA signals given in ppm.

Docking calculations for complexes of **1** and **2** with DNA were performed with AutoDock 4.2[47] with a 64 points box centered in the A/T DNA region (spacing 0.35 Å) and using a genetic algorithm. The coordinates of the crystal structure of Dickerson dodecamer (PDB[49] ID 1PRP) were employed for the calculation after refinement with Sybyl computational program. The ligands were built and minimized using Macromodel from Schrodinger suite software.

4.6. References

(1) Boehr, D. D.; Nussinov, R.; Wright, P. E. *Nat. Chem. Biol.* **2009**, *5* (12), 954.
(2) Tompa, P. *FEBS Lett.* **2005**, *579* (15), 3346.
(3) Tompa, P. *Nat. Chem. Biol.* **2012**, *8* (7), 597.
(4) Dunker, A. K.; Romero, P.; Obradovic, Z.; Garner, E. C.; Brown, C. J. *Genome Informatics* **2000**, *11*, 161.
(5) Wright, P. E.; Dyson, H. J. *J. Mol. Biol.* **1999**, *293* (2), 321.
(6) Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradović, Z.; Dunker, A. K. *J. Mol. Biol.* **2002**, *323* (3), 573.
(7) Tompa, P. *BioEssays* **2003**, *25* (9), 847.
(8) Bussell, R.; Eliezer, D. *J. Biol. Chem.* **2001**, *276* (49), 45996.
(9) Gunasekaran, K.; Tsai, C.-J.; Kumar, S.; Zanuy, D.; Nussinov, R. *Trends Biochem. Sci.* **2003**, *28* (2), 81.
(10) Fuxreiter, M.; Simon, I.; Friedrich, P.; Tompa, P. *J. Mol. Biol.* **2004**, *338* (5), 1015.
(11) Uversky, V. N. *Eur. J. Biochem.* **2002**, *269* (1), 2.
(12) Wright, P. E.; Dyson, H. J. *Curr. Opin. Struct. Biol.* **2009**, *19* (1), 31.
(13) Vucetic, S.; Brown, C. J.; Dunker, A. K.; Obradovic, Z. *Proteins Struct. Funct. Bioinforma.* **2003**, *52* (4), 573.
(14) Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradović, Z. *Biochemistry* **2002**, *41* (21), 6573.

(15) Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. *J. Mol. Biol.* **2004**, *337* (3), 635.
(16) TOMPA, P.; CSERMELY, P. *FASEB J.* **2004**, *18* (11), 1169.
(17) Kosol, S.; Contreras-Martos, S.; Cedeño, C.; Tompa, P. *Molecules*. **2013**, pp 10802–10828.
(18) Palmer, A. G. *Chem. Rev.* **2004**, *104* (8), 3623.
(19) Miller, M.; Shuman, J. D.; Sebastian, T.; Dauter, Z.; Johnson, P. F. *J. Biol. Chem.* **2003**, *278* (17), 15178.
(20) Oyadomari, S.; Mori, M. *Cell Death Differ.* **2004**, *11* (4), 381.
(21) Singh, V. K.; Pacheco, I.; Uversky, V. N.; Smith, S. P.; MacLeod, R. J.; Jia, Z. *J. Mol. Biol.* **2008**, *380* (2), 313.
(22) Tabas, I.; Ron, D. *Nat. Cell Biol.* **2011**, *13* (3), 184.
(23) Kaufman, R. J. *J. Clin. Invest.* **2002**, *110* (10), 1389.
(24) Wishart, D. S.; Sykes, B. D. *J. Biomol. NMR* **1994**, *4* (2), 171.
(25) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104* (17), 4546.
(26) Hwang, T. L.; Mori, S.; Shaka, A. J.; Van Zijl, P. C. M. *J. Am. Chem. Soc.* **1997**, *119* (26), 6203.
(27) Hwang, T. L.; van Zijl, P. C. M.; Mori, S. *J. Biomol. NMR* **1998**, *11* (2), 221.
(28) Bollag, D. M.; McQueney, P. A.; Zhu, J.; Hensens, O.; Koupal, L.; Liesch, J.; Goetz, M.; Lazarides, E.; Woods, C. M. *Cancer Res.* **1995**, *55* (11), 2325 LP.
(29) Kowalski, R. J.; Giannakakous, P.; Hamel, E. *J. Biol. Chem.* **1997**, *272* (4), 2534.
(30) Kamath, K.; Jordan, M. A. *Cancer Res.* **2003**, *63* (18), 6026 LP.
(31) Prota, A. E.; Bargsten, K.; Zurwerra, D.; Field, J. J.; Díaz, J. F.; Altmann, K.-H.; Steinmetz, M. O. *Science* **2013**, *339* (6119), 587 LP.
(32) Nogales, E.; Whittaker, M.; Milligan, R. A.; Downing, K. H. *Cell* **1999**, *96* (1), 79.
(33) Snyder, J. P.; Nettles, J. H.; Cornett, B.; Downing, K. H.; Nogales, E. *Proc. Natl. Acad. Sci.* **2001**, *98* (9), 5312.
(34) Díaz, J. F.; Barasoain, I.; Andreu, J. M. *J. Biol. Chem.* **2003**, *278* (10), 8407.
(35) Buey, R. M.; Calvo, E.; Barasoain, I.; Pineda, O.; Edler, M. C.; Matesanz, R.; Cerezo, G.; Vanderwal, C. D.; Day, B. W.; Sorensen, E. J.; Lopez, J. A.; Andreu, J. M.; Hamel, E.; Diaz, J. F. *Nat Chem Biol* **2007**, *3* (2), 117.
(36) Coley, H. M. *Cancer Treat. Rev.* **2008**, *34* (4), 378.
(37) Beer, T. M.; Smith, D. C.; Hussain, A.; Alonso, M.; Wang, J.; Giurescu, M.; Roth, K.; Wang, Y. *Br. J. Cancer* **2012**, *107* (5), 808.
(38) McMeekin, S.; Patel, R.; Verschraegen, C.; Celano, P.; Burke II, J.; Plaxe, S.; Ghatage, P.; Giurescu, M.; Stredder, C.; Wang, Y.; Schmelter, T. *Br J Cancer* **2012**, *106* (1), 70.
(39) Bystricky, B.; Chau, I. *Expert Opin. Investig. Drugs* **2011**, *20* (1), 107.
(40) Lechleider, R. J.; Kaminskas, E.; Jiang, X.; Aziz, R.; Bullock, J.; Kasliwal, R.; Harapanhalli, R.; Pope, S.; Sridhara, R.; Leighton, J.; Booth, B.; Dagher, R.; Justice, R.; Pazdur, R. *Clin. Cancer Res.* **2008**, *14* (14), 4378 LP.
(41) Gaugaz, F. Z.; Redondo-Horcajo, M.; Barasoain, I.; Díaz, J. F.; Cobos-Correa, A.; Kaufmann, M.; Altmann, K. H. *ChemMedChem* **2014**, *9* (10), 2227.
(42) Kuzniewski, C. N.; Gertsch, J.; Wartmann, M.; Altmann, K. H. *Org. Lett.* **2008**, *10* (6), 1183.
(43) Nicolaou, K. C.; Namoto, K.; Ritze, A.; Ulven, T.; Shoji, M.; Li, J.; Amico, G. D.; Liotta, D.; French, C. T.; Wartmann, M. *Analysis* **2001**, 3, 9313.
(44) Field, J. J.; Pera, B.; Calvo, E.; Canales, A.; Zurwerra, D.; Trigili, C.; Rodríguez-Salarichs, J.; Matesanz, R.; Kanakkanthara, A.; Wakefield, S. J.; Singh, A. J.; Jiménez-Barbero, J.; Northcote, P.; Miller, J. H.; López, J. A.; Hamel, E.; Barasoain, I.; Altmann, K. H.; Díaz, J. F. *Chem. Biol.* **2012**, *19* (6), 686.
(45) Canales, A.; Rodríguez-Salarichs, J.; Trigili, C.; Nieto, L.; Coderch, C.; Andreu, J. M.; Paterson, I.; Jiménez-Barbero, J.; Díaz, J. F. *ACS Chem. Biol.* **2011**, *6* (8), 789.

(46) Canales, A.; Nieto, L.; Rodríguez-Salarichs, J.; Sánchez-Murcia, P. A.; Coderch, C.; Cortés-Cabrera, A.; Paterson, I.; Carlomagno, T.; Gago, F.; Andreu, J. M.; Altmann, K. H.; Jiménez-Barbero, J.; Díaz, J. F. *ACS Chem. Biol.* **2014**, *9* (4), 1033.
(47) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. *J. Comput. Chem.* **2009**, *30* (16), 2785.
(48) Nettles, J. H.; Li, H.; Cornett, B.; Krahn, J. M.; Snyder, J. P.; Downing, K. H. *Science (80-.).* **2004**, *305* (5685), 866 LP.
(49) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28* (1), 235.
(50) Lehn, J.-M. *Chem. – A Eur. J.* **1999**, *5* (9), 2455.
(51) Rowan, S. J.; Cantrill, S. J.; Cousins, G. R. L.; Sanders, J. K. M.; Stoddart, J. F. *Dynamic Covalent Chemistry*; **2002**; Vol. 41.
(52) Ramström, O.; Lehn, J.-M. *Nat. Rev. Drug Discov.* **2002**, *1* (1), 26.
(53) Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *Drug Discov. Today* **2002**, *7* (2), 117.
(54) Congreve, M. S.; Davis, D. J.; Devine, L.; Granata, C.; O'Reilly, M.; Wyatt, P. G.; Jhoti, H. *Angew. Chemie Int. Ed.* **2003**, *42* (37), 4479.
(55) Li, J.; Nowak, P.; Otto, S. *J Am Chem Soc* **2013**, *135* (25), 9222.
(56) Ramström, O.; Bunyapaiboonsri, T.; Lohmann, S.; Lehn, J.-M. *Biochim. Biophys. Acta - Gen. Subj.* **2002**, *1572* (2), 178.
(57) Mondal, M.; Hirsch, A. K. H. *Chem. Soc. Rev.* **2015**, *44* (8), 2455.
(58) Herrmann, A. *Chem. Soc. Rev.* **2014**, *43* (6), 1899.
(59) Mondal, M.; Radeva, N.; Fanlo-Virgós, H.; Otto, S.; Klebe, G.; Hirsch, A. K. H. *Angew. Chemie Int. Ed.* **2016**, *55* (32), 9422.
(60) Cousins, G. R. L.; Poulsen, S.-A.; Sanders, J. K. M. *Chem. Commun.* **1999**, No. 16, 1575.
(61) Furlan, R. L. E.; Ng, Y.-F.; Otto, S.; Sanders, J. K. M. *J. Am. Chem. Soc.* **2001**, *123* (36), 8876.
(62) Liu, J.; West, K. R.; Bondy, C. R.; Sanders, J. K. *Org. Biomol. Chem.* **2007**, *5* (5), 778.
(63) Bunyapaiboonsri, T.; Ramström, O.; Lohmann, S.; Lehn, J.-M.; Peng, L.; Goeldner, M. *ChemBioChem* **2001**, *2* (6), 438.
(64) Clipson, A. J.; Bhat, V. T.; McNae, I.; Caniard, A. M.; Campopiano, D. J.; Greaney, M. F. *Chem. - A Eur. J.* **2012**, *18* (34), 10562.
(65) Bhat, V. T.; Caniard, A. M.; Luksch, T.; Brenk, R.; Campopiano, D. J.; Greaney, M. F. *Nat. Chem.* **2010**, *2* (6), 490.
(66) Caraballo, R.; Dong, H.; Ribeiro, J. P.; Jiménez-Barbero, J.; Ramström, O. *Angew. Chem. Int. Ed.* **2010**, *49* (3), 589.
(67) Orphanides, G.; Reinberg, D. *Cell* **2002**, *108* (4), 439.
(68) Darnell, J. E. J. *Nat. Rev. Cancer* **2002**, *2* (10), 740.
(69) Look, A. T. *Science (80-.).* **1997**, *278* (5340), 1059.
(70) Brennan, P.; Donev, R.; Hewamana, S. *Mol. Biosyst.* **2008**, *4* (9), 909.
(71) Vogt, P. K. *Oncogene* **2001**, *20* (19), 2365.
(72) Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387* (6629), 202.
(73) Vazquez, M. E.; Caamano, A. M.; Mascarenas, J. L. *Chem. Soc. Rev.* **2003**, *32* (6), 338.
(74) Ren, J.; Jenkins, T. C.; Chaires, J. B. *Biochemistry* **2000**, *39* (29), 8439.
(75) Haworth, I. S.; Elcock, A. H.; Freeman, J.; Rodger, A.; Richards, W. G. *J. Biomol. Struct. Dyn.* **1991**, *9* (1), 23.
(76) Hamilton, P. L.; Arya, D. P. *Nat. Prod. Rep.* **2012**, *29* (2), 134.
(77) Haq, I.; Ladbury, J. E.; Chowdhry, B. Z.; Jenkins, T. C.; Chaires, J. B. *J. Mol. Biol.* **1997**, *271* (2), 244.
(78) Mazur, S.; Tanious, F. A.; Ding, D.; Kumar, A.; Boykin, D. W.; Simpson, I. J.; Neidle, S.; Wilson, W. D. *J. Mol. Biol.* **2000**, *300* (2), 321.
(79) Haq, I. *Arch. Biochem. Biophys.* **2002**, *403* (1), 1.

(80) Neidle, S. *Nat. Prod. Rep.* **2001**, *18* (3), 291.
(81) Dickerson, R. E.; Drew, H. R. *J. Mol. Biol.* **1981**, *149* (4), 761.
(82) Greenidge, P. A.; Jenkins, T. C.; Neidle, S. *Mol. Pharmacol.* **1993**, *43* (6).
(83) Lane, a N.; Jenkins, T. C.; Brown, T.; Neidle, S. *Biochemistry* **1991**, *30* (5), 1372.
(84) Jenkins, T. C.; Lane, A. N.; Neidle, S.; Brown, D. G. *Eur. J. Biochem.* **1993**, *213* (3).
(85) Sánchez, M. I.; Vázquez, O.; Martínez-Costas, J.; Vázquez, M. E.; Mascareñas, J. L. *Chem. Sci.* **2012**, *3* (7), 2383.

Chapter 5

Conclusions

Different molecular recognition processes of biochemical interest have been characterized by the application of NMR.

-The molecular recognition features of Tn-antigen mimetics have been revealed towards different lectins from different sources. The overall picture of tricyclic compound **1** reveals can mimic the Tn-antigen with a similar binding mode as Tn reference model (α-OMeGalNAc) displaying higher affinity than Tn-reference for the case of ECL whilst MGL affinity is lower than the reference. On the other hand, glycopeptide mimetic **2** was also recognized by lectins with a similar binding mode to that observed for mimetic **1**. All these data permit to consider the tricyclic mimetic **1** as a rigid but selective mimic of the Tn antigen suitable for further biomedical applications, always depending on the chemical architecture of the receptor binding site.

-The multivalency of a bifunctional ligand targeting different proteins, hgal-3 and MMP12, was proved. The binding mode with each of the proteins was assessed by STD-NMR as well as the amino acids from the binding site of each protein were elucidated by ^{1}H-^{15}N HSQC (receptor point of view methodology). Evidences of simultaneous binding to both proteins were achieved by NMR. The spectroscopic data suggest that bifunctional ligand is able to target both proteins simultaneously, terminal lactose epitope is recognized by hgal-3 and hydroxamic acid by MMP12.

-Three different monoclonal anti-glycan antibodies targeting the xenoantigen α-Gal were tested with a xenoantigen disaccharide model applying ligand point of view NMR experiments, STD-NMR, proving that they are suitable receptors for the xeno-epitope. However, different affinities, qualitatively estimated by STD intensities, suggest that anti-gal A4 antibody displays the strongest affinity and may be considered as possible candidate for further biomedical applications.

-The binding of two anti-sialic antibodies was examined by STD-NMR displaying a complete selectivity towards compared to Neu5Ac. Moreover, these monoclonal antibodies exhibited preference towards the α-anomer of NeuGc confirmed by the recognition of NeuGc α-linked trisaccharides.

-Enzymatical synthesis of sialylated glycans of different lengths was achieved for further molecular recognition studies. The linear synthesized glycans were shown, by STD-NMR, to display different affinity towards hemagglutinins (HA) of different influenza strains through the terminal Neu5Ac moiety and depending of the regiochemistry of Neu5Ac linkage. By means of

docking calculations a 3D model of the binding mode of the glycans was proposed, in accordance with the spectroscopic data and displaying an umbrella-like topology for the case of Neu5Acα2-6Gal linked glycans. Furthermore, very low affinity was observed when Neu5Acα2-6Gal linked glycans were tested against HAs from avian strains specific of Neu5Acα2-3Gal linked glycans showing up the importance of local mutations for influenza transfection between species.

- The NMR dynamic analysis (^{1}H-^{15}N hetNOE and R_2/R_1 relaxation ratio) of a fragment of 53 amino acids in the N-terminal region of CHOP has revealed evidences of structural order in the Ser59 to Glu65. The obtained experimental NMR data (positive hetNOE and large R_2/R_1 ratios) signature of of structured protein domains together with a previous chemical shift index analysis, indicate that this region may adopt a α-helix motif. Furthermore, ^{1}H-^{2}H chemical exchange experiments (CLEANEX-PM) confirmed the structuring propensity of that segment displaying a significantly reduced chemical exchange with water solvent in comparison with the rest of the peptide.

- The molecular recognition of a cyclopropane epothilone derivative (EPO-tc) by tubulin (in both, dimeric and microtubule states of aggregation) was characterized by combined NMR and molecular-modelling approach. The STD-derived epitopes and the obtained bound conformations in both states were similar to those described for natural epothilone-A (EPO-A) with no significant changes in the binding modes. Therefore, the experimental data suggest that EPO-tc may be an efficient epothilone derivative with improved properties and a potential candidate for anti-tumor therapy.

- NMR methodology (STD-NMR, DOSY and tr-NOESY) has been proved to be a versatile tool to direct detection of binders in dynamic combinatorial chemistry (DCC). Using a simplified aldehyde/hydrazine dynamic combinatorial library (DCL), it was possible to reveal that hydrazone **3** is bound and amplified by the frequenin protein, Frq2, in the presence of hydrazone **5** validating the role of Frq2 as a selective template. Furthermore, STD-based competition studies permitted to assess that the amplified product **3** competes for the same binding site than a well-known Frq2-binder, chlorpromazine (CPZ).

- The interaction of synthetic DNA-minor groove binders has been studied by NMR. The specificity towards A/T-rich regions of DNA of ligands **1** and **2** and structure of their complexes were characterized by NOESY experiments. The combination of the obtained NMR data with

docking calculation have permitted to propose a 3D model of the interaction revealing that both ligands establish hydrogen bonding in the minor groove of a dsDNA helix.